



INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

18905

I A. R. I. 6.

S. C. P-1/8/47-P. J.-17-5-48 2000

PROCEEDINGS
OF THE
ROYAL SOCIETY OF LONDON

SERIES B—BIOLOGICAL SCIENCES

VOL 131

LONDON

**Printed and published for the Royal Society
By the Cambridge University Press
Bentley House, N.W. 1**

2 November 1943

PRINTED IN GREAT BRITAIN BY
WALTER LEWIS, M.A.
AT THE CAMBRIDGE UNIVERSITY PRESS

CONTENTS

SERIES B VOLUME 131

No. B 862. 14 September 1942

	PAGE
The starling (<i>Sturnus vulgaris</i> L.) and foot-and-mouth disease. By W. S. Bullough	1
The physiology of incompatibility in plants. I. The effect of temperature. By D. Lewis. (Plate 1)	13
Some experiments on the trichromatic theory of vision. By H. V. Walters	27
The selection of invisible mutations. By K. Mather and L. G. Wigan	50
The integument and moult cycle of <i>Tegenaria atrica</i> (Araneae). By H. C. Browning	65
The development of the bristles in normal and some mutant types of <i>Drosophila melanogaster</i> . By A. D. Lees and C. H. Waddington	87

No. B 863. 10 December 1942

Croonian Lecture. Chromatic behaviour. By L. Hogben, F.R.S.	111
Bakerian Lecture. Amino-acid analysis and the structure of proteins. By A. C. Chibnall, F.R.S.	136
Evolution in plants by kaleidoscopic mutation. By J. C. Willis, F.R.S.	161
The production of cancer by pure hydrocarbons. VI. By G. M. Badger, J. W. Cook, F.R.S., C. L. Hewett, E. L. Kennaway, F.R.S., N. M. Kennaway and R. H. Martin	170
Blood sedimentation: a study in haemo-mechanics. By R. B. Whittington	183

No. B 864. 14 May 1943

Anniversary Meeting of the Royal Society and Newton Tercentenary Celebrations. (7 Plates)	191
Ferrier Lecture. Fatigue following highly skilled work. By F. C. Bartlett, F.R.S.	247
The effects of cold and colchicine on mitosis in the newt. By H. N. Barber and H. G. Callan. (Plates 9-11)	258
The visceral nervous system of the earthworm: I. Nerves controlling the tone of the alimentary canal. By N. Millott	271

No. B 865. 2 November 1943

PAGE

Pilgrim Trust Lecture. Organization of American scientists for the war. By K. T. Compton	297
✓The fate of haemoglobin in <i>Rhodnius prolixus</i> (Hemiptera) and other blood- sucking arthropods. By V. B. Wigglesworth, F.R.S.	313
The spectral sensitivity of the fovea and extrafovea in the Purkinje range. By H. V. Walters and W. D. Wright	340
The visceral nervous system of the earthworm. II. Evidence of chemical trans- mission and the action of sympatheticomimetic and parasympatheticomimetic drugs on the tone of the alimentary canal. By N. Millott	362
The origin of the primitive tetrapod limb. By T. S. Westoll	373
The hyomandibular of <i>Eusthenopteron</i> and the tetrapod middle ear. By T. S. Westoll	393
Index	415

ERRATUM

Proc. Roy. Soc. B, volume 130, number 861

On page 343 in figure 4 the percentages given as ordinates
have been entered as 0.1 to 0.9; they should be 1 to 9.

The starling (*Sturnus vulgaris* L.) and foot-and-mouth disease

BY W. S. BULLOUGH, B.Sc., PH.D.

Department of Zoology, University of Leeds

(Communicated by J. Hammond, F.R.S.—Received 17 November 1941)

Circumstantial evidence is recorded which suggests the possibility of a connexion between the starling (*Sturnus vulgaris* L.) and the occurrence of unexplained outbreaks of foot-and-mouth disease. The starling is singled out as the one common species of bird in the British Isles which is intimately connected with cattle, and it is shown that striking correlations exist between the migrations, movements, and distribution of this species and the geographical and monthly incidence of the disease.

An attempt is also made to analyse the position of foot-and-mouth disease in Finland, Sweden, and Denmark. In the case of Sweden, where the starling is absent in autumn and winter, a graph is obtained which is the reverse of that for the British Isles where, in autumn and winter, the starling is most numerous.

1. INTRODUCTION

Attempts have often been made to trace some connexion between the habits and migrations of wild animals and the incidence in the British Isles of the foot-and-mouth disease of cattle, and farmers commonly make a habit of blaming an abnormal prevalence of one animal or another when outbreaks occur in their neighbourhood. Throughout the annual *Reports of Proceedings under the Diseases of Animals Acts*, published by the Ministry of Agriculture and Fisheries, repeated reference is made to some local abundance of wild animals, and among others, rats, ground game, rooks, gulls, and starlings have been specifically mentioned. The possible relation of wild mammals to foot-and-mouth disease has been studied for many years by the Oxford Bureau of Animal Population, and it is possible that such animals may hold a natural reservoir of the disease. It is known that cats, dogs, hedgehogs, and field voles are susceptible, and in most of these species it has been shown that the virus may pass from one animal to another by contact infection (*Fifth Progress Report*). Hedgehogs and rats infected with the virus have been found on infected premises (Longley 1937). Such animals, however, although they may hold the virus, cannot by themselves carry the disease very far, and they are almost incapable of bringing it into the British Isles from the continent of Europe. Ticks and insects have also been suspected of acting as carriers, but there is little evidence for this, and it is doubtful whether they too could carry the disease for any great distance.

In 1923, Stockman & Garnett made a serious attempt to connect the migrations of birds, known to take place between Europe and the British Isles, with the monthly incidence of the disease. Their work was open to several criticisms, some of which were made from an ornithological standpoint by Thomson (1924). However, they did come to the valuable conclusion that the only obvious way in

which infection can be carried over long distances, often in a very short time, is through the agency of birds. One difficulty is that those birds which have been tested do not easily contract the disease, and that even when the virus does succeed in establishing itself, it does not spread further by contact alone (*Fifth Progress Report*). The virus, however, is so easily transmitted by mechanical means that the contraction of the disease by the bird itself is not a vital factor. Birds which visit infected pastures or premises may easily contaminate their feet or feathers, or they may pick up the virus when feeding and pass it unharmed through their alimentary canals. It is known that the virus may retain its viability for as long as two or three months on such dead materials as feeding stuffs or bedding material. A bird whose body was once infected would therefore retain for some time the power to transmit the disease, and in that time it might travel a very considerable distance. If on migration, it might come from the continent into the British Isles.

This paper is concerned only with the possible methods of transmission of the disease. It does not attempt, except in an incidental manner, to suggest a solution to the problem of its origin. Further, it is only concerned with those outbreaks of foot-and-mouth disease which, it is stated in the *Reports*, cannot be explained, or which, it is suggested, may have been due to the movement of birds. The great majority of outbreaks are easily explicable on the grounds of direct contact between cattle, or of the infection of foodstuffs or bedding, or of the movements of people or vehicles, but the mysterious outbreaks are the initial ones from which the others are derived by some such direct contact. These initial unexplained outbreaks may occur almost anywhere, and at any time of the year. They may be separated from the nearest infected centre by many hundreds of miles, or they may be separated from the last outbreak by many months. It has been found, however, that they tend to be more frequent in some areas of the British Isles than others, and more prevalent in some months of the year. During a course of research into the breeding cycles and general biology of the starlings (*Sturnus vulgaris* L.), which are either resident in this country or which visit it in autumn and winter, several points arose which suggested the existence of a possible connexion between the normal habits, movements, and distribution of this species and the incidence of foot-and-mouth disease in the British Isles and Scandinavia, and in the various months of the year.

2. OBSERVATIONS

(a) *Association of the starling with cattle*

The starling is the one common species of British bird which is intimately associated with pastureland, with farms, and with cattle. The bird is almost omnivorous, but for preference it feeds on insects or molluscs. Starlings, keeping closely together in flocks, excavate their food from the pastures, and if possible, they prefer to feed in the immediate vicinity of cattle or sheep. The birds are frequently seen feeding within inches of the noses and feet of the grazing animals,

and they often perch on their backs. These habits have been frequently commented on but never fully explained. In winter time, when the ground is either frozen or covered with snow, starlings flock in the vicinity of houses and farms, and groups of them enter the cattle sheds, perch on the backs of the animals and on the feeding troughs, and pick up any grain or other edible matter which they can find. There are many other species of birds, such as the skylark or the peewit, which frequent pastureland, and there are insectivorous species, such as the pied wagtail, which often feed close to the cattle and even perch on them. The starling is, however, the only species which is present in such enormous numbers, which keeps together in such large flocks, and which, for preference, feeds around or even on the cattle themselves whether outdoors or indoors.

Having thus separated the starling as the bird most intimately connected with cattle and therefore most likely to carry any disease, several lines of inquiry are opened.

(b) Starling migrations and the incidence of the disease

The starlings seen in the British Isles are derived from two sources. First, there are the resident birds which are mainly sedentary (Bullough 1942*a*), and second, there are the birds of continental origin which, coming from Scandinavia and the neighbouring countries, spend most of the autumn and winter in this country. The continental birds arrive here in late September, October, and early November, and leave again for their nesting places in March and sometimes early April. These facts have been established by the marking of starlings which has been done on a large scale both in this country and on the continent. The great majority of the records of the journeys made by continental starlings to and from the British Isles, since ringing was begun nearly forty years ago, are to be found in past volumes of *British Birds* (Witherby & Leach). The detailed analysis of these results has produced some interesting facts. Birds of the continental race have been ringed in their native countries and recovered in the British Isles, and birds ringed in the British Isles in autumn and winter have been recovered on the continent in spring and summer. Of ten birds coming from or going to Norway, seven were reported in Scotland, and only one in Northern Ireland and two in England. Of 563 birds coming from or going to continental countries other than Norway, 559 were reported in England, Wales, and Ireland, and only four in Scotland. Thus, although the figures for Norway are unfortunately very small, it would appear that the majority of the migrating Norwegian birds come to Scotland, and that the majority of the other continental starlings come to England, Wales, and Ireland. It appears probable that the Norwegian birds follow a northern migration route across the North Sea, perhaps via the Orkney and Shetland Islands, whereas the other continental birds, unable to cross the Swedish and Norwegian mountains, come down the coasts of the Baltic, cross the Low Countries, and enter England via the east coast (figure 1). From other evidence two similar general migration routes have been described by Thomason (1926).

This conclusion is significant because foot-and-mouth disease is rare in Scotland and almost non-existent in Norway, whereas it is common in England and Wales and very common in those countries, such as Sweden, Denmark, Germany, and Holland, whence come the greater number of the starlings which arrive each autumn in East Anglia. Even in such a bad year as 1925 when in Denmark, for instance, there were 53,567 outbreaks of foot-and-mouth disease (Petersen 1939) and in Sweden 4,664 outbreaks (Magnusson 1939), Norway was entirely free (Abel 1939). For comparison in this same year there were 232 outbreaks (from all causes) in England and Wales, but Scotland remained free. Any abnormally

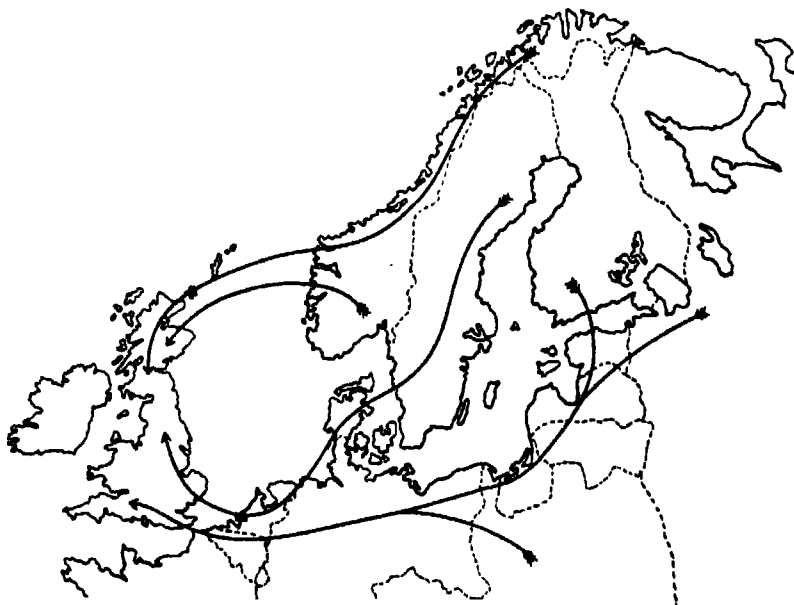


FIGURE 1. Probable migration routes to the British Isles of starlings from Norway and other continental countries.

large outbreak of the disease on the continent usually spreads to England and Wales, and it tends to appear first in autumn at migration time. For instance, in 1925, when Europe was experiencing one of the worst epidemics, eighteen unexplained outbreaks of the disease occurred in England and Wales in October, November, and December. This compares with nine such outbreaks in all the other months of that year. The disease again became common in the autumn of 1926, when the epidemic was still raging on the continent, and during another epidemic the same phenomenon was apparent in the autumn months of 1930 and 1932.

(c) *Starling roosts and the incidence of the disease*

An indication has been given of a possible way in which the virus of foot-and-mouth disease might be introduced into the British Isles from the continent of Europe. The evidence that the starling transmits the disease from place to place

within this country has now to be considered, and from this point of view the bird has one habit of primary importance. For seven months of the year (June to December) almost all British starlings leave their nesting localities each evening to roost communally in some wood or reedbed, and during the whole of the time (October to March) that the continental starlings are present in the British Isles, they also travel each day between their feeding grounds and these same communal roosting places. The journeys made are considerable, and a bird often flies a total of 40 or 50 miles each day for this purpose alone. During these daily movements the virus of foot-and-mouth disease might be carried to the roost by some infected bird and there passed to other birds which, next morning, might take it away in many other directions. Opportunities for the spread of the virus in the roost at night are very great because large numbers of sleeping birds are packed closely together shoulder to shoulder. At certain seasons of the year, one roost often contains between fifty thousand and a hundred thousand starlings each night. The numbers of birds present in the communal roosts in each month of the year provide an index of the numbers travelling daily across the country, and therefore they may also provide an index of the probability that the virus will be spread. Repeated counts were made at communal roosts over a period of nearly two years, and although unfortunately no one roost was occupied for a whole period of twelve months, the main seasonal variations in the numbers of roosting birds were worked out. In all, eight roosts were discovered, and the counts were made as the birds arrived at night. With practice, these counts proved very much easier to make than appeared at first sight. In all cases, two or three observers counted at the same time, and the different results so obtained were found to approximate very closely. An average was taken of these different results, and in some cases the final figure was checked by photographic evidence. The figures in table 1 were obtained at two roosts, each of which was occupied for a different period of the year, and allowing for the difference of season, these roosts, one near Beamsley in Wharfedale and the other at North Deighton near Wetherby, were similar in size. In the table an average of the counts made in each month is given. In the two months September 1939 and February 1940, the roosts became deserted, and it was found that such changes to new roosting places frequently occur. In each of these months, therefore, instead of taking the average of the counts made, the maximum figure recorded is taken as being representative of the month. When, owing usually to adverse weather conditions, the original counts were only approximate, the prefix *c.* (*circa*) is inserted before the average figure, and because of the probable error in the case of the larger figures, these are expressed to the nearest five thousand birds. The main changes recorded in this table were confirmed by a study of the variations in the numbers of birds at the other six roosts which were discovered, and by a comparison with the observations made by Campbell (1902) on a big roost in Cramond Island in the Firth of Forth.

As the figures given in table 1 are an index of the numbers of starlings travelling backwards and forwards across the British Isles in each month of the year, and

therefore of the probability that the virus of foot-and-mouth disease may be spread, they may be compared with the monthly totals of unexplained outbreaks of this cattle disease. The figures given in table 2 are the total numbers of such outbreaks in the British Isles from 1900 to 1937 inclusive. The figures for the years 1900-21 are taken from Stockman & Garnett (1923), and those for the succeeding years have been compiled from the annual *Reports of Proceedings under the Diseases of Animals Acts*. As already stated, only those outbreaks are included which could not be explained or which were supposed to have been due to birds. The comparison between tables 1 and 2 is made in figures 2 and 3, and it is seen how closely they coincide. There is a rise in September and October, and both graphs remain high throughout the autumn. The increase in the numbers of roosting birds at this time is directly due to the arrival, between the end of September and the beginning of November, of millions of starlings from the continent, and the rise

TABLE 1. APPROXIMATE AVERAGE NUMBERS (IN THOUSANDS) OF STARLINGS OCCUPYING COMMUNAL ROOSTS IN THE DIFFERENT MONTHS OF THE YEAR

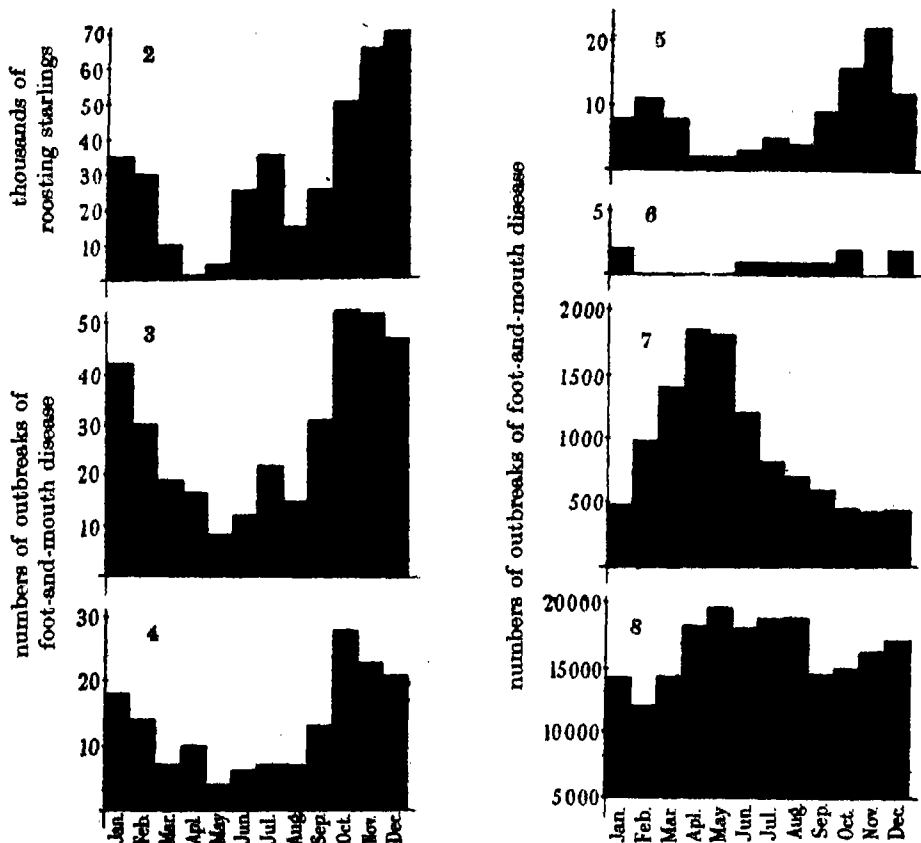
Beamsley roost								North Deighton roost			
1939								1940			
Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.
c. 10	1	4	25	35	15	25	50	c. 65	70	c. 35	c. 30

TABLE 2. NUMBERS OF UNEXPLAINED OUTBREAKS OF FOOT-AND-MOUTH DISEASE IN THE BRITISH ISLES BETWEEN 1900 AND 1937

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
42	30	19	16	8	12	22	15	31	54	53	47

in the graph of foot-and-mouth disease may also be connected with this and with the fact that such large numbers of birds, both British and continental, are moving backwards and forwards across the British Isles each day. From December to February the fall in both graphs may be correlated with the fact that, at the end of December and in January, most adult British starlings cease to occupy the communal roosts, and remain each night to sleep in their nesting holes (Bullough 1942a). The numbers of birds which move across the country each day are, in this way, roughly halved. It may also be that, by this time, preventative measures have succeeded in stamping out much of the disease which was introduced in the autumn. In March the continental birds leave this country and return to their nesting places. The roosts then reach their minimum size, and throughout April they are mainly occupied by first year and non-breeding males (Bullough 1942a). The roosts remain small in May when foot-and-mouth disease is at a minimum. In June the starlings and their newly fledged young return to sleep communally, and in July the roosts are again very populous. Large numbers of birds are moving across the country each day, and the graph of foot-and-mouth disease rises to a small peak in July. In August both graphs fall, and in the case of the

starling roosts, this is due to the yearly moult of the flight feathers. The power of flight is seriously weakened, and in exceptional cases it may even be entirely lost (Bullough 1942b). Long flights to the roosts are not possible, especially if the weather is wet. The effect of this was found to affect different roosts in different degrees. Both graphs rise again in September when, with the flight feathers fully



FIGURES 2-8

FIGURE 2. Average numbers of starlings occupying communal roosts in the different months of the year.

FIGURE 3. Numbers of unexplained outbreaks of foot-and-mouth disease in the British Isles between 1900 and 1937.

FIGURE 4. Numbers of unexplained outbreaks of foot-and-mouth disease in the eastern area of England between 1922 and 1937.

FIGURE 5. Numbers of unexplained outbreaks of foot-and-mouth disease in the western area of England and in Wales between 1922 and 1937.

FIGURE 6. Numbers of unexplained outbreaks of foot-and-mouth disease in Scotland between 1922 and 1937.

FIGURE 7. Total numbers of outbreaks of foot-and-mouth disease in Sweden between 1900 and 1937.

FIGURE 8. Total numbers of outbreaks of foot-and-mouth disease in Denmark between 1900 and 1937.

formed, the British birds again make long daily flights, and when the first continental birds begin to arrive.

It was considered possible that if the starlings coming in autumn from the various continental countries (other than Norway) arrived first on the east coast of England, and only later spread right across the country, the peak of the monthly incidence of foot-and-mouth disease might be earlier in the eastern than in the western counties. A line (formed by the western boundaries of Northumberland, Durham, Yorkshire, Nottinghamshire, Leicestershire, Northamptonshire, Buckinghamshire, Surrey, and Hampshire) was drawn down the centre of England dividing it into eastern and western parts. Wales was included in the western part. From 1922 to 1937 inclusive, there were 131 unexplained outbreaks of foot-and-mouth disease in the eastern area, and 102 unexplained outbreaks in the western area. The numbers of outbreaks in each month of the year in these two areas are shown in table 3. The numbers of outbreaks in Scotland during this period are also given, but as there were only ten of these, their significance, if any, cannot be assessed.

TABLE 3. NUMBERS OF UNEXPLAINED OUTBREAKS OF FOOT-AND-MOUTH DISEASE IN THE EASTERN, WESTERN, AND SCOTTISH AREAS BETWEEN 1922 AND 1937

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Eastern area											
18	14	7	10	4	6	7	7	13	28	23	21
Western area											
8	11	8	2	2	3	5	4	9	16	22	12
Scottish area											
2	0	0	0	0	1	1	1	1	2	0	2

The comparison between the figures for these different areas is shown in the figures 4, 5, and 6, and it is seen that the greatest number of outbreaks in the eastern area occurs in October, whereas in the western area the peak is not reached until November. It is possible that this difference may be due to the time taken by the starlings to cross the country, and in this connexion it is interesting to note that whereas in the western area the outbreaks reach a minimum in April, in the eastern area there is a small peak in that month. This might possibly be connected with the return journey across the eastern area made at the end of March by the starlings from the western area.

(d) Incidence of the disease on the continent

It was hoped, if the starling should prove to be the main carrier of foot-and-mouth disease, that by studying the monthly incidence of the disease in some northern European country, where the bird is present in spring and summer but not in autumn and winter, a graph the reverse of that shown for the British Isles might be obtained. It was found, however, that such northern countries as Finland are almost free from the disease (Kankaanpää, 1939), and it appears that some

unknown factor or factors are responsible for its suppression in these latitudes. The most northern country in which the disease is relatively common, and for which statistics are available, is Sweden, and this country is especially suitable for study as the starling is absent in autumn and winter. The numbers of outbreaks of foot-and-mouth disease in each month of the year, from 1900 to 1937 inclusive, were obtained from Professor H. Magnusson, and these figures are given below in table 4. Unfortunately, as no distinction is made between primary unexplained outbreaks and secondary outbreaks due to known causes, the figures refer to the total number of outbreaks. They are therefore not strictly comparable with the figures already given for the British Isles, and only broad conclusions may be drawn from them. These figures are expressed in figure 7, and it is clear that this graph is a direct opposite to that for the British Isles. There is a sharp rise in the incidence of the disease in February and March, and the peak for the year is reached in April. This coincides with the return of the birds from the south. Again, instead of a rise in autumn, the graph falls to a minimum in November and December. Because of the inclusion of the secondary outbreaks, no more detailed correlations than these can be noted.

TABLE 4. TOTAL NUMBERS OF OUTBREAKS OF FOOT-AND-MOUTH DISEASE IN SWEDEN BETWEEN 1900 AND 1937

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
489	975	1396	1855	1810	1203	823	696	609	464	426	427

TABLE 5. TOTAL NUMBERS OF OUTBREAKS OF FOOT-AND-MOUTH DISEASE IN DENMARK BETWEEN 1900 AND 1937

Jan.	Feb.	Mar.	Apr.	May	June
14,202	11,957	14,223	18,330	19,616	18,094
July	Aug.	Sept.	Oct.	Nov.	Dec.
18,784	18,789	14,505	14,908	16,112	17,120

Denmark is another country for which figures of the monthly incidence of foot-and-mouth disease are available (Petersen, 1939), and it forms an interesting contrast to both Finland and Sweden. Starlings are present in Denmark all the year round, and the country is crossed by the migration route of the starlings from the Baltic countries. Foot-and-mouth disease is extremely prevalent, and as in the case of Sweden, the figures given in table 5 include all outbreaks, explained and unexplained, from 1900 to 1937 inclusive. As the disease is so prevalent, its ramifications must be extremely involved, and the number of secondary outbreaks must be high. Figure 8 is an expression of these figures, and it is evident that the disease is very common all through the year. As in Sweden, however, there is a sharp rise in late winter and early spring which may possibly be connected with the northward migration of starlings from countries to the south and west where the disease is common. There is no comparable rise in autumn, and perhaps this

may be connected with the fact that most of the birds then crossing Denmark come from northern countries which are relatively free from foot-and-mouth disease. Any attempted correlation of this graph with the movements of starlings is, however, difficult, and it appears certain that there are unknown factors present in Denmark which are favourable to the presence and spread of the disease at all times of the year. Some of these factors may well be the opposite of those which so effectively eliminate the disease from Finland.

(e) *Distribution of the starling and of foot-and-mouth disease*

One further piece of evidence, the comparison of the distributions in the British Isles of starlings and of foot-and-mouth disease, may be mentioned. This comparison must necessarily be confined mainly to England and Wales, as the lack of foot-and-mouth disease in Scotland, where starlings are common, has already been described. It was shown by Marples (1934), who made a census of the large starling roosts present in England, Wales, and Scotland in the autumn and winter of 1932-3, that such roosts are not commonly found above the 600 ft. level. Of 285 roosts discovered, only five were above that altitude, and Marples concluded that this was due to the fact that the starling habitually feeds in pastureland. On the average, above 600 ft. there are only rough hill pastures or heather moors, and for preference these habitats are avoided. However, it occasionally happens, especially in the Pennines between Yorkshire and Lancashire, that both towns and cultivation are found as high as 1000 ft., and where this is so, the starling follows too. Marples' figure of 600 ft. constitutes an average, and an attempt was made to find whether the unexplained initial outbreaks of foot-and-mouth disease also occurred, on the average, below 600 ft. Figure 9 shows the distribution of starling roosts and of unexplained outbreaks of foot-and-mouth disease. The outbreaks charted are those occurring in the years 1922-37 inclusive, and it is seen, except in odd cases, that the outbreaks of foot-and-mouth disease are in fact found below 600 ft. Of 270 outbreaks only two were above that level. Of course, the majority of the cattle are also found below the 600 ft. level, but on all except the highest hills there is a large population of hill sheep and, in some places, of deer which appear to be almost entirely immune. Many of these animals are brought down to lower ground in the winter, and therefore before the possible significance of this correlation between the distributions of starlings and of foot-and-mouth disease can be assessed, it will be necessary to find the proportions of sheep and cattle above and below the 600 ft. level at all times of the year.

3. CONCLUSIONS

The observations and correlations here set forward constitute the most circumstantial of evidence. The starling, like most birds, probably does not contract foot-and-mouth disease at all easily, and it is suggested that it may merely act as a mechanical carrier. This makes absolute proof of the present thesis extremely

difficult to obtain, and the difficulty is not eased by the fact that only one bird in several millions needs to be infected to produce the results noted. It is also quite certain that many other factors, such as that connected with latitude, must enter

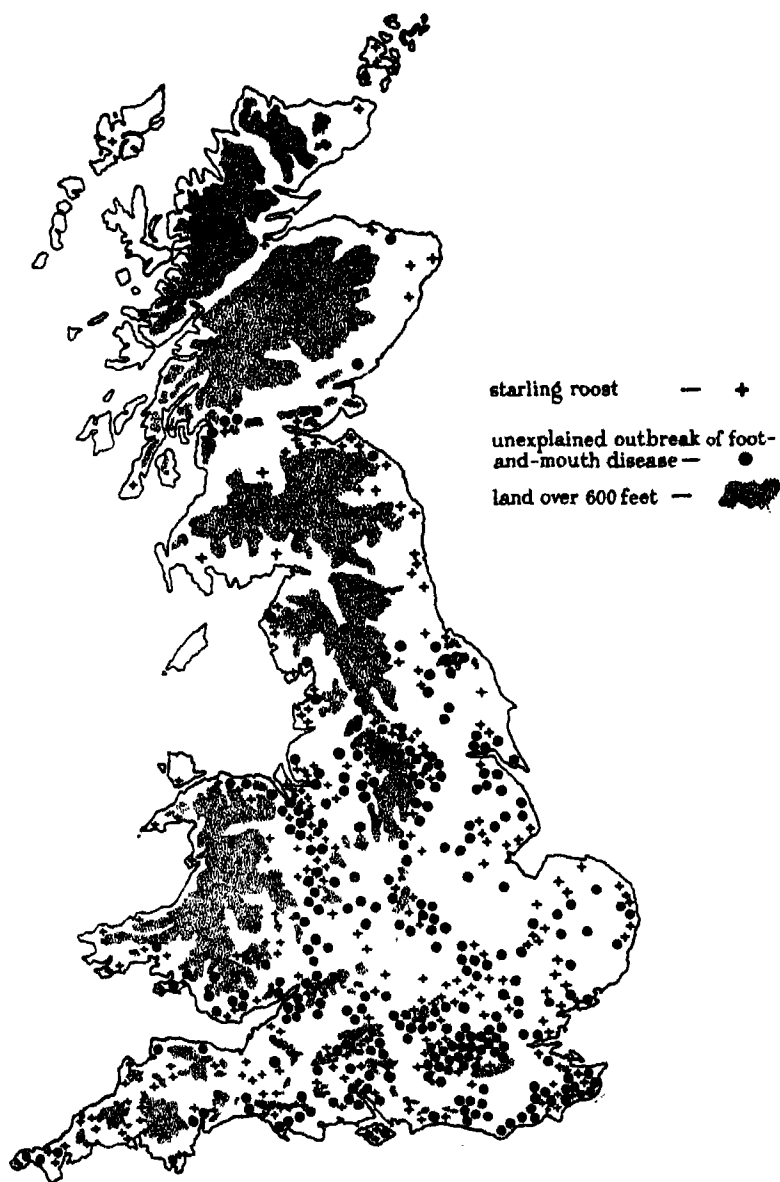


FIGURE 9. Distribution of starling roosts and of unexplained outbreaks of foot-and-mouth disease.

into the problem. The weather may exert some seasonal effect, other species of birds may be involved to a lesser degree, and it is possible that in some species of wild mammal there exists a natural reservoir of the disease. However, as the

suggestions put forward in this paper cannot, at least for some time, be carried any further, it is hoped that they may prove helpful to others engaged in this line of research.

The author is very much indebted to Dr and Mrs R. Carrick for their help in field work and for their valuable criticisms, to Miss M. Denton for her great assistance in the preparation of this paper, and to Professor E. A. Spaul who revised the manuscript. Grateful acknowledgements are also due to Dr V. Kan-kaanpää, Maatalousministeriö, Helsinki, to Herr A. Abel, Veterinærdirektøren, Oslo, to Professor H. Magnusson, Veterinærbakteriologiska Laboratoriet, Malmö, and to Herr G. W. Petersen, Veterinærdirektoratet, Copenhagen, for the detailed reports which they sent of the incidence of foot-and-mouth disease in their respective countries.

REFERENCES

- Abel, A. 1939 Private communication.
 Bullough, W. S. 1942a *Phil. Trans. B*, **231**, 165.
 Bullough, W. S. 1942b *Ibis*, Series 14, vol. 6, p. 225.
 Campbell, C. 1902 *Ann. Scot. Nat. Hist.* p. 2.
 Kankeanpää, V. 1939 Private communication.
 Longley, E. O. 1937 *Fifth Progress Report of the Foot-and-Mouth Disease Research Committee*, p. 34.
 Magnusson, H. 1939 Private communication.
 Marples, B. J. 1934 *J. Anim. Ecol.* **3**, 187.
 Petersen, G. W. 1939 Private communication.
 Stockman, S. & Garnett, M. 1923 *J. Minist. Agric.* **30**, 681.
 Thomson, A. L. 1924 *Nature, Lond.*, **113**, 52.
 Thomson, A. L. 1926 *Problems of bird migration*, p. 49. London.
 Witherby, H. F. & Leach, E. P. 1931-40 *British Birds*, **25**, 110, 357; **26**, 352; **28**, 106; **29**, 132; **31**, 14; **33**, 62.
Fifth Progress Report of the Foot-and-Mouth Disease Research Committee. Ministry of Agriculture and Fisheries, London.
Reports of Proceedings under the Diseases of Animals Acts for the Years 1922 to 1937 inclusive. Ministry of Agriculture and Fisheries. London.

The physiology of incompatibility in plants

I. The effect of temperature

By D. LEWIS

John Innes Horticultural Institution, Merton, London

(Communicated by C. D. Darlington, F.R.S.—Received 20 January 1942)

[Plate 1]

Two different genetic systems of incompatibility between pollen and style are known. One, heterogamety, depends upon the genotype of the individual pollen grain; the other, heterostyly, upon the genotype of its parent. We do not know whether the two types are physiologically related. The specificity of heterogamety indicates an immunity reaction. The effect of temperature on pollen-tube growth in the two systems was measured in order to discover their relationship.

Compatible pollinations of both systems showed increased rate of growth with increased temperature until the lethal point was approached at about 35° C.

Incompatible pollinations of both systems showed an optimum growth rate between 15 and 20° C. The physiological method of inhibition is therefore probably related although its genetic basis is different.

The different rate of growth at different temperatures gives different total growth at inhibition, and at the most favourable temperature there may be no inhibition at all. There is therefore no specific inhibitory zone in these plants, although elsewhere the top of the style may provide such a zone.

Certain genotypes of *Oenothera organensis* show such powerful incompatibility that no temperature sensitivity can be discovered. This extreme modification is determined by the pollen parent's genotype, like the main action in heterostyly.

In heterostyled plants thrum pollen has to grow down the longer pin style; it is adapted to this in two ways. In *Primula* it is larger, in *Linum grandiflorum* it has a higher osmotic pressure. In either case, presumably, it has the higher dry weight.

In two heterostyled *Primula* species thrum pollen grows faster down the long-pin style than pin pollen does down the short-thrum style. But in the illegitimate matings thrum pollen is more strongly inhibited. There is therefore a differentiation of the mechanism adapted to secure equal regularity of cross-fertilization of the two types.

1. INTRODUCTION

Self-incompatibility in plants can be classified genetically according to whether the incompatibility reaction of the pollen is determined (1) by its gametic constitution, or (2) by the constitution of the maternal tissue from which it arose. The gametic determination in its simplest and most common form is by a series of multiple allelomorphs designated $S_1, \dots, S_k$. Pollen grains cannot effect fertilization in a flower which has a style carrying the same sterility allelomorph, so that self-fertilization is impossible (East & Mangelsdorf 1925). Modifications of this type of determination are occasionally found, such as two sets of sterility allelomorphs in autotetraploids (Lawrence 1930) and in *Abutilon* (Sears 1937).

Maternal determination of self-incompatibility is found commonly in species with distylic or tristylic flowers, and occasionally in homostyled plants. In distylic plants, one pair of genes T, t is concerned. The short-styled plant (thrum)

suggestions put forward in this paper cannot, at least for some time, be carried any further, it is hoped that they may prove helpful to others engaged in this line of research.

The author is very much indebted to Dr and Mrs R. Carrick for their help in field work and for their valuable criticisms, to Miss M. Denton for her great assistance in the preparation of this paper, and to Professor E. A. Spaul who revised the manuscript. Grateful acknowledgements are also due to Dr V. Kankaanpää, Maatalousministeriö, Helsinki, to Herr A. Abel, Veterinærdirektøren, Oslo, to Professor H. Magnusson, Veterinærbakteriologiska Laboratoriet, Malmö, and to Herr G. W. Petersen, Veterinærdirektoratet, Copenhagen, for the detailed reports which they sent of the incidence of foot-and-mouth disease in their respective countries.

REFERENCES

- Abel, A. 1939 Private communication.
 Bullough, W. S. 1942a *Phil. Trans. B*, **231**, 165.
 Bullough, W. S. 1942b *Ibis*, Series 14, vol. 6, p. 225.
 Campbell, C. 1902 *Ann. Scot. Nat. Hist.* p. 2.
 Kankaanpää, V. 1939 Private communication.
 Longley, E. O. 1937 *Fifth Progress Report of the Foot-and-Mouth Disease Research Committee*, p. 34.
 Magnusson, H. 1939 Private communication.
 Marples, B. J. 1934 *J. Anim. Ecol.* **3**, 187.
 Petersen, G. W. 1939 Private communication.
 Stockman, S. & Garnett, M. 1923 *J. Minist. Agric.* **30**, 681.
 Thomson, A. L. 1924 *Nature, Lond.*, **113**, 52.
 Thomson, A. L. 1926 *Problems of bird migration*, p. 49. London.
 Witherby, H. F. & Leach, E. P. 1931-40 *British Birds*, **25**, 110, 357; **26**, 352; **28**, 106; **29**, 132; **31**, 14; **33**, 62.
Fifth Progress Report of the Foot-and-Mouth Disease Research Committee. Ministry of Agriculture and Fisheries, London.
Reports of Proceedings under the Diseases of Animals Acts for the Years 1922 to 1937 inclusive. Ministry of Agriculture and Fisheries. London.

The physiology of incompatibility in plants

I. The effect of temperature

BY D. LEWIS

John Innes Horticultural Institution, Merton, London

(Communicated by C. D. Darlington, F.R.S.—Received 20 January 1942)

[Plate 1]

Two different genetic systems of incompatibility between pollen and style are known. One, heterogamety, depends upon the genotype of the individual pollen grain; the other, heterostyly, upon the genotype of its parent. We do not know whether the two types are physiologically related. The specificity of heterogamety indicates an immunity reaction. The effect of temperature on pollen-tube growth in the two systems was measured in order to discover their relationship.

Compatible pollinations of both systems showed increased rate of growth with increased temperature until the lethal point was approached at about 35° C.

Incompatible pollinations of both systems showed an optimum growth rate between 15 and 20° C. The physiological method of inhibition is therefore probably related although its genetic basis is different.

The different rate of growth at different temperatures gives different total growth at inhibition, and at the most favourable temperature there may be no inhibition at all. There is therefore no specific inhibitory zone in these plants, although elsewhere the top of the style may provide such a zone.

Certain genotypes of *Oenothera organensis* show such powerful incompatibility that no temperature sensitivity can be discovered. This extreme modification is determined by the pollen parent's genotype, like the main action in heterostyly.

In heterostyled plants thrum pollen has to grow down the longer pin style; it is adapted to this in two ways. In *Primula* it is larger, in *Linum grandiflorum* it has a higher osmotic pressure. In either case, presumably, it has the higher dry weight.

In two heterostyled *Primula* species thrum pollen grows faster down the long-pin style than pin pollen does down the short-thrum style. But in the illegitimate matings thrum pollen is more strongly inhibited. There is therefore a differentiation of the mechanism adapted to secure equal regularity of cross-fertilization of the two types.

1. INTRODUCTION

Self-incompatibility in plants can be classified genetically according to whether the incompatibility reaction of the pollen is determined (1) by its gametic constitution, or (2) by the constitution of the maternal tissue from which it arose. The *gametic* determination in its simplest and most common form is by a series of multiple allelomorphs designated $S_1, \dots, S_k$. Pollen grains cannot effect fertilization in a flower which has a style carrying the same sterility allelomorph, so that self-fertilization is impossible (East & Mangelsdorf 1925). Modifications of this type of determination are occasionally found, such as two sets of sterility allelomorphs in autotetraploids (Lawrence 1930) and in *Abutilon* (Sears 1937).

Maternal determination of self-incompatibility is found commonly in species with distylic or tristylic flowers, and occasionally in homostyled plants. In distylic plants, one pair of genes T, t is concerned. The short-styled plant (thrum)

is *Tt* and the long-styled plant (pin) is *tt*. Pollen from a thrum plant fails to effect adequate fertilization on a thrum style, irrespective of whether the actual pollen grain is carrying *T* or *t*, but this pollen is fully effective on a pin style. Thus the behaviour of the pollen depends solely upon the genetic constitution of the mother plant from which it came, and not on its own genetic constitution. In some trimorphic heterostyled plants and in *Capsella grandiflora* two pairs of genes similar in their action to the *T*, *t* genes of the dimorphic plants are involved, one pair being epistatic to the other.

Self-incompatibility, irrespective of the genetic determination, is due either to the inhibition of pollen-tube growth at some stage between germination and entry into the embryo-sac, or to inhibition of fertilization (Sears 1937). The genetic specificity of this inhibition indicates that self-incompatibility is the result of a reaction of the immunity type between the growing pollen tube and the stylar tissue.

Certain conditions are effective in breaking down self-incompatibility either completely or partially, as follows:

Bud pollination	} <i>Nicotiana</i> (East 1923)
End of flowering season pollination	
Mutilation of stigma	} <i>Petunia</i> (Yasuda 1930)
Keeping plant in the dark	
	<i>Brassica oleracea</i> (Kakizaki 1930)
	<i>Oenothera organensis</i> (Emerson 1940)

However, these effects are not common to all self-incompatible plants. Temperature differences should have a similar effect on all self-incompatible plants in which the mechanism is an immunity reaction. To test this, the following series of measurements of pollen-tube growth were made.

2. TECHNIQUE

Pollinations in all species studied were made on flowers cut from the plant and kept in water. Before extensive experiments were made on any species, the effect on pollen-tube growth of cutting the flower from the plant was examined, and in no case was there a significant difference. The flowers were transferred, 30 min. before pollination, to incubators thermostatically controlled to $\pm 0.5^{\circ}\text{C}$. The styles were fixed in 3 parts absolute alcohol to 1 part glacial acetic acid. I am indebted to Miss I. Modibowska for assistance in the preparation and pollination of the styles.

The preparation of the style for examination varied according to the species. The treatment in *Oenothera organensis* was the same as that used by Emerson (1938), i.e. slitting the cortex, pulling out the strand of conducting tissue and staining and mounting in cotton blue. For *Primula obconica* and *P. sinensis* the styles were cut on the freezing microtome at 20μ thickness or longitudinally into two halves by hand. The sections were stained and mounted in a hot mixture of acid fuchsin and light green. The styles of *Prunus avium* were cut on the freezing microtome and stained in cotton blue.

3. *OENOTHERA ORGANENSIS*

Self-incompatibility in *Oenothera organensis* is determined by the gametic constitution of the pollen. Its inheritance is by a series of S allelomorphs similar to those first found in *Nicotiana*, and since found in the majority of self-sterile plants. The differentiation of pollen-tube growth of incompatible and compatible pollen at normal temperatures in the style is very marked. Incompatible pollen tubes stop growing after 1 hr. and do not usually grow more than 1–2 mm. Compatible pollen tubes have a uniform rate of growth and reach the ovary, 160–180 mm. below the stigma, in approximately 24 hr. (Emerson 1938). Emerson carried out his pollen-tube growth tests at 25° C, a temperature at which incompatible tubes are strongly inhibited.

The material used in the present study was obtained originally from Dr Emerson, and the sterility allelomorphs referred to are the same as those used by him.

The effect of temperature on compatible pollen-tube growth was studied in reciprocal crosses between an S_2S_6 and an S_3S_6 plant. Measurements of the longest pollen tube in the style were made at 4½ and 24 hr. after pollination (figure 1). The length of the longest tube was used as an estimate of pollen-tube growth, because this length is correlated with the mean length of the one class of tube. Furthermore, in the cross $S_2S_6 \times S_3S_6$ there are two classes of tubes, incompatible and compatible, therefore the mean length of all pollen tubes in the style cannot be taken as an estimate of compatible tube growth.

The growth rate of compatible pollen tubes increases with an increase in temperature to a maximum at 30–33° C and then falls off rapidly as the lethal temperature is approached. This kind of growth-temperature curve is similar to that in the self-fertile *Datura Stramonium* (Buchholz & Blakeslee 1927). In both compatible pollinations in a self-sterile plant and all pollinations in a self-fertile plant, the effect of temperature is typical of that for most biological growth rates and is simply the result of increased absorption and metabolism at higher temperatures.

Incompatible pollen-tube growth at different temperatures was studied in self-pollinations and in crosses between genetically different plants with the same sterility allelomorphs. The results are given in tables 1 and 2 and summarized in figures 1 and 2. Instead of the pollen-tube growth rate increasing with increasing temperature, there is a decrease in the pollen-tube growth rate at higher temperatures. The maximum pollen-tube growth is at 15° C, and at this temperature the growth for the first 4½ hr. is as good as that of compatible tubes. At 20° C there is in general a slower rate of growth than at 15° C, but some plants show a very much greater retardation of pollen-tube growth than others. Pollen tubes in all plants at 25 and 30° C are very strongly inhibited after 1–2 mm. growth.

The growth curves show that the effect of high temperature is to decrease the growth rate during the whole period of growth, rather than to bring about complete inhibition earlier (figure 2). This agrees with the conclusions of Emerson

(1940), based on style grafting experiments, that there is no particular inhibitory zone in the style or inhibitory substance which diffuses up the style from the ovary, as claimed in *Petunia* by Yasuda (1934).

TABLE 1. POLLEN-TUBE GROWTH IN INCOMPATIBLE SELF-POLLINATIONS AT DIFFERENT TEMPERATURES IN *OENOTHERA ORGANENSIS*. THE MEAN LENGTHS ARE THE MEANS OF MAXIMUM TUBE LENGTHS. THE NUMBER OF FLOWERS ON WHICH EACH MEAN IS BASED IS INDICATED BY THE FIGURE IN BRACKETS (cf. figure 1)

plant	hr.	mean lengths				
		15° C	20° C	25° C	30° C	33° C
1 ¹ (S ₂ S ₆)	4½	1.0 (5)	1.0 (3)	1.0 (3)	1.0 (3)	—
1 ¹⁴ (S ₂ S ₆)	4½	15.5 (2)	11.0 (1)	1.4 (4)	1.3 (2)	—
1 ¹⁸ (S ₂ S ₆)	4½	12.0 (6)	10.2 (5)	1.0 (2)	1.0 (7)	—
2 ¹⁷ (S ₄ S ₆)	4½	13.2 (4)	4.6 (3)	1.5 (2)	1.7 (7)	1.5 (2)
1 ¹⁸ (S ₃ S ₆)	24	54.5 (2)	22.0 (1)	2.5 (2)	2.0 (3)	—
2 ¹⁷ (S ₄ S ₆)	24	30.0 (1)	25.6 (3)	12.0 (1)	1.5 (2)	—

TABLE 2. VARIATION IN SAMPLES OF MORE THAN FIVE INDIVIDUALS (cf. table 1)

1 ¹⁸	4½ hr.	15°	7,	8,	12,	15,	15,	15,	—
1 ¹⁸	4½ hr.	20°	7,	10,	10,	11,	13,	—,	—
1 ¹⁴	4½ hr.	25°	1,	1,	1,	2,	2,	—,	—
2 ¹⁷	4½ hr.	30°	1,	1,	1,	2,	2,	3,	3

Thus if there were an inhibitory zone, the tubes would grow more rapidly at higher temperatures until they reached this zone and they would then cease to grow. Similarly, if inhibitory substances diffused up from the ovary after the flower opened, until these substances reached the pollen tubes there would be a period of more rapid growth at high temperatures, although the onset of inhibition may take place earlier owing to the more rapid diffusion of the substances. Further evidence on this point was obtained by keeping flowers at different temperatures for 12 hr. just before pollination. Flowers of 2¹⁷ (S₄S₆) were kept at 15 and 25° C for 12 hr., self-pollinated and placed in 15, 20 and 25° C. There was no effect of the pre-treatment on pollen-tube growth. If the incompatibility is due to a substance which diffuses up the style, the temperature before pollination should have an effect.

In some strains of *Oenothera organensis* the incompatibility reaction is so strong that even at low temperatures the pollen tubes are inhibited (figure 1). This is best shown by the great difference between plant 1¹ and all the other plants tested. Plant 1¹ has such a strong incompatibility reaction that it never has self-pollen tubes longer than 1–2 mm. at any temperature; it is therefore *temperature insensitive*. Plant 1², on the other hand, has pollen tubes no longer than those of 1¹ at 25° C, but at 15° C has tubes 12 mm. long after 4½ hr.; it is therefore *temperature sensitive*. Plant 1¹ and 1² are both S₂S₆. The ability of the self-pollen tubes to grow at low temperatures is therefore not due to a gene at the S locus.



FIGURE 10. Styles of pin flowers of *Primula obconica* 24 hr. after pollination: (a) At 20° C, maximum length of pollen tubes = 1.4 mm., entering the narrow part of the style. (b) At 30° C, maximum length of pollen tubes = 0.4 mm. The swollen ends of the tubes can be seen below the stigma ($\times 140$).

The progeny of crosses between plants with long self-pollen tubes at 15° C (temperature sensitive) and plant 1¹ (temperature insensitive) were tested for their self-incompatibility allelomorphs and their sensitivity to low temperature (table 3). The two types were easily distinguished; the temperature insensitive (Mm) had tubes 1-2 mm. long, while the temperature sensitive had tubes 8-12 mm. long. Two plants which were doubtful, with 4 mm. growth, were not included. Plants

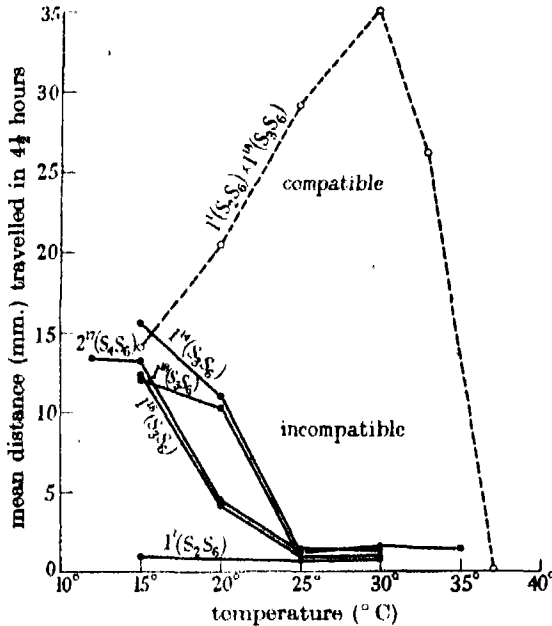


FIGURE 1

FIGURE 1. Length of pollen tubes 4½ hr. after pollination at different temperatures, showing the different effect of temperature between incompatible and compatible tubes in *Oenothera organensis*. The incompatible pollinations vary in their reaction to low temperatures according to the type of plant selfed, as shown by the difference between the incompatible curves (---- compatible pollination, ——— incompatible pollination).

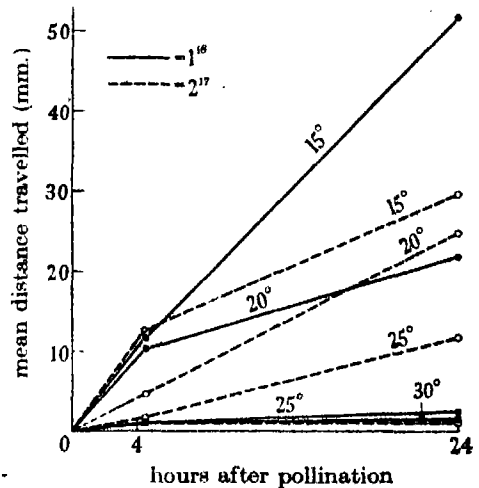


FIGURE 2

FIGURE 2. Growth curves of incompatible tubes in *Oenothera organensis* at different temperatures, showing that the effect of high temperature is to decrease the growth rate during the whole period of growth.

segregated for temperature sensitivity in about equal numbers, and there was a slight indication of linkage between the incompatibility S locus and the sensitivity-modifying gene m. It seems therefore that the modifying effect is mainly due to a single-gene difference. In *Nicotiana* the situation seems to be different. East (1934) considered that the strength of incompatibility varied mainly between different S allelomorphs. Some determined fertility with selfing in the bud stage and others did not.

Emerson (1940) found, as I have, that the degree of self-pollen-tube growth varied between plants both with different and the same S allelomorphs. Emerson also

found some plants with greater variability of self-pollen-tube growth than others. It now seems that these differences in variability are due to differences in the temperature sensitivity, as shown above.* A plant similar to 1^1 being unaffected by temperature will always give extremely uniform pollen-tube growth, while a plant such as 1^2 gives variable results unless the environment is rigidly controlled.

TABLE 3. SEGREGATION OF S GENES AND MODIFYING GENES IN
OENOTHERA ORGANENSIS

parents	progeny			
	S_2S_4 , t.s.	S_2S_4 , t.i.	S_6S_4 , t.s.	S_6S_4 , t.i.
$1^1 (S_1S_6)$, t.i. $\times$ $2^{17} (S_6S_4)$, t.s.	4	2	2	3
$1^1 (S_2S_6)$, t.i. $\times$ $1^{18} (S_2S_6)$, t.s.	S_2S_3 , t.s. 4	S_2S_3 , t.i. 1	S_2S_6 , t.s. 2	S_2S_6 , t.i. 2

TABLE 4. MEAN POLLEN-TUBE GROWTH $4\frac{1}{2}$ HR. AFTER POLLINATION OF 1^1 AND 1^2 SELFED AND OF THE RECIPROCAL CROSSES BETWEEN THEM, SHOWING THAT THE DIFFERENCE IN TEMPERATURE SENSITIVITY IS IN THE POLLEN AND NOT IN THE STYLE (*OENOTHERA ORGANENSIS*)

temp. °C	$1^1 (S_2S_6)$ selfed t.i.	$1^2 \times 1^1$	$1^2 (S_2S_6)$ selfed t.s.	$1^1 \times 1^2$
15-16	1.0	1.0	12.3	9.6
20	1.0	1.0	4.5	1.0
30	1.0	1.0	1.0	1.0

Pollen-tube growth of reciprocal pollinations between $1^2 (S_2S_6mm)$ and $1^1 (S_2S_6Mm)$ were studied to determine how the modifying gene *M* works. Must it be in the pollen, in the styles or in both? The results in table 4 show that it is only necessary for the pollen parent to carry the modifying gene. Furthermore, its action depends upon the constitution of the maternal tissue from which the pollen came, and not on the gametic constitution of the pollen, since two kinds of tubes in a style were not found.

4. *PRUNUS AVIUM*

All varieties of the sweet cherry, *Prunus avium*, are self-incompatible, and their incompatibility is due to a series of multiple allelomorphs like those of *Oenothera organensis* (Crane & Lawrence 1931; Crane & Brown 1937). Also as in *Oenothera*, incompatibility in the cherry is a very rigorous mechanism; approximately one fruit sets for every thousand flowers pollinated. Furthermore, incompatibility is due to the inhibition of pollen tubes half-way down the style (Roy 1939).

* Emerson (1938) states that, in his experiments, the flowers were kept in a moist chamber at 25° C, but elsewhere in the same paper he states that the temperature ranged from 20 to 27° C.

Compatible pollen-tube growth at different temperatures was studied in the cross Guigne d'Annonay $\times$ Bedford Prolific (figure 3). Just as in *Oe. organensis*, the growth rate increased as the temperature increased until the lethal point was reached. However, the optimum temperature for pollen-tube growth is 25° C in *Prunus* as compared with 33° C in *Oenothera*. This difference shows that natural selection has been acting on the pollen tubes. *Prunus avium*, occupying temperate regions and flowering in the spring, would be accustomed to temperatures ranging from 5 to 20° C, while *Oenothera*, growing in the hot regions of Arizona, is subjected to much higher temperatures.

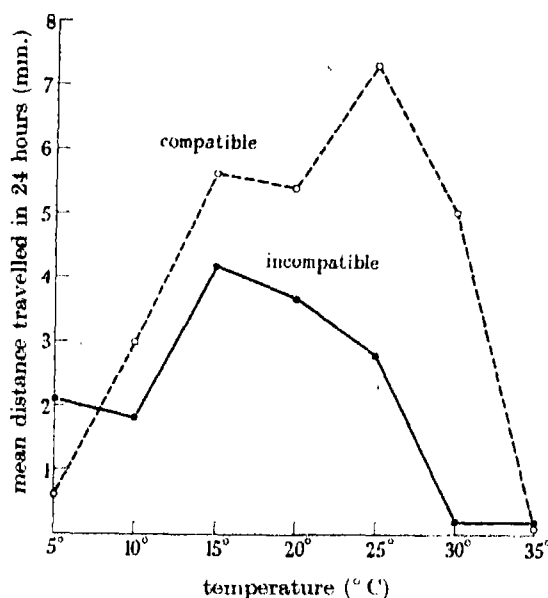


FIGURE 3

FIGURE 3. Length of pollen tubes at different temperatures in *Prunus avium* (----- compatible cross-pollination, ——— incompatible self-pollination).

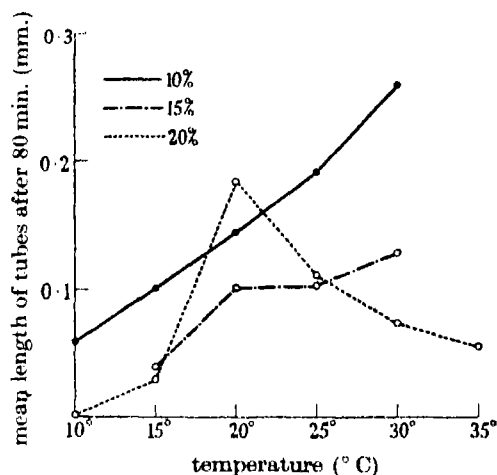


FIGURE 4

FIGURE 4. Pollen-tube growth in varying concentrations of sugar solution at different temperatures in *Prunus avium*.

Pollen of Bedford Prolific was germinated in hanging drops of sugar solution of different concentrations and at different temperatures. The results of pollen-tube measurements under these conditions, which are given in figure 4, give some information on the difference in the optimum temperature found between species. In 10 and 15 % sugar solution the optimum temperature for growth is 30° C, but in 20 % sugar the optimum is 20° C. In the style the optimum temperature is 25° C. Thus by a decrease in effective osmotic pressure of the style the optimum temperature can be raised.

Incompatible pollen-tube growth of Bedford Prolific selfed at different temperatures shows a maximum growth at 15° C and a rapid fall in growth rate at

higher temperatures (figure 3 and table 5). This again agrees with the *Oenothera* results. At temperatures below 15° C the growth rate is approximately the same as that of the compatible tubes.

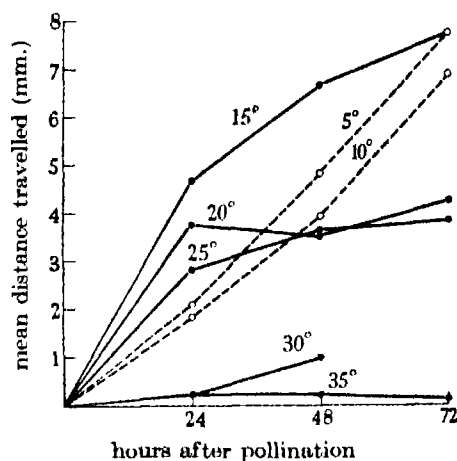


FIGURE 5. Growth curves of incompatible tubes in *Prunus avium*.

TABLE 5. MEAN POLLEN-TUBE GROWTH IN *PRUNUS AVIUM* VAR. BEDFORD PROLIFIC SELFED AT DIFFERENT TEMPERATURES. THE FIGURES IN BRACKETS INDICATE THE NUMBER OF FLOWERS

temp. °C	mean lengths		
	24 hr.	48 hr.	72 hr.
5	2.1 (4)	4.8 (1)	7.7 (1)
10	1.8 (3)	3.9 (1)	6.9 (2)
15	4.6 (3)	6.6 (2)	7.6 (1)
20	3.7 (3)	3.3 (1)	4.3 (1)
25	2.8 (1)	3.7 (1)	3.8 (1)
30	0.2 (1)	1.0 (1)	—
35	0.2 (1)	0.2 (1)	0.1 (1)

The temperature-growth curve of the incompatible tubes thus seems to be the result of two opposing actions, (a) the increase in growth rate due to increased metabolism and absorption at higher temperatures, which is the predominant effect below 15° C, and (b) the increase in the rate of the incompatibility reaction at higher temperatures, which is predominant above 15° C. The corresponding curve in *Oe. organensis* does not include temperatures below 12° C, consequently only the second effect is apparent.

The growth curves of incompatible pollen tubes over different periods of time after pollination fall into three groups: (figure 5) (a) the 5, 10, and 15° C, (b) the 20 and 25° C, and (c) the 30 and 35° C curves. The lower temperature curves (group a) show a slow but fairly constant rate of growth up to at least 72 hr. The second group (b) curves also show a similar growth rate during the first 24 hr.,

but very little growth takes place after this time. The higher temperature curves (group c) show very little growth during the first 24 hr. and little after this time. Thus the incompatible tubes, unlike those in *Oenothera*, have a period of steady growth and then are arrested at some point in the style.

The sudden arrest of growth has been found in other self-sterile plants and has usually been interpreted as due to the presence of a localized zone in the style which is a block to pollen-tube growth. In *Prunus avium* the position of arrest can be changed by altering the temperature. Thus at a medium temperature the tubes cease growing about half-way down the style; at low temperatures about three-quarters down, and at high temperatures already in the stigmatic lobe. Therefore instead of there being a zone of inhibition, it is probable that there is a mutual reaction between the pollen tube and the whole of the stylar tissue, as in *Oenothera*. The sudden arrest of growth is then due to the accumulation of the effect of the incompatibility reaction to a value beyond which growth stops.

5. *PRIMULA OBCONICA*

Self-sterility in the plants so far described is due to a reaction between pollen and style, the reaction of the pollen being determined by its gametic constitution. *Primula obconica* is a distylic plant, and as in most distylic plants the reaction of the pollen is not determined by its gametic constitution, but by the constitution of the maternal tissue from which it arose. The two types of flowers are *pin* (long style, short anthers) and *thrum* (short style, long anthers). Illegitimate matings, i.e. *pin* × *pin* or *thrum* × *thrum*, are almost completely sterile. This high degree of sterility has made it difficult to determine the inheritance of distyly; however, in all other species examined *thrum* is dominant to *pin* and there is no reason to believe that the same is not true in *P. obconica*.

The results of pollen-tube growth measurements are given in table 6 and figure 6. Pollen-tube growth of legitimate, i.e. compatible matings, increases with increasing temperature to the optimum, which is 30° C, and then falls rapidly to 35° C.

Pollen-tube growth of illegitimate, i.e. incompatible, matings is not only very much slower than that of legitimate matings but has an optimum temperature at 20° C. Thus the degree of pollen-tube growth of the two types of matings and the effect of temperature (figure 10, plate 1) are similar to those in *Oenothera* and *Prunus*.

High temperature might increase the inhibition of illegitimate pollen tubes by directly accelerating the reaction between pollen tube and style, or by increasing the amount of an inhibitory substance in the style prior to any reaction. To test this, flowers were kept at different temperatures before pollination. After pollination the flowers were kept at 20° C, the optimum temperature for illegitimate pollen-tube growth. Table 7 gives the results. Although slight differences occur between the pretreatments, probably due to secondary effects such as differential development of the flower, it is clear that there is no effect comparable with the effect of

TABLE 6. *PRIMULA OBCONICA*, MAXIMUM SELFED POLLEN-TUBE GROWTH AT DIFFERENT TEMPERATURES IN MM.

temp. °C	plant 1 (thrum)		plant 5 (thrum)		plant 2 (pin)		plant 3 (pin)		plant 4 (pin)	
	hr.		hr.		hr.		hr.		hr.	
	24		24	48	24	48	24	48	24	48
10	0.1		0.1	0.1	0.1	0.2	0.1	0.1	0.3	0.3
15	0.2		0.1	0.0	0.1	0.5	0.2	0.0	0.9	—
20	0.7		0.3	0.3	1.2	3.4	1.4	1.7	0.6	0.9
25	0.3		0.0	0.0	0.1	0.1	0.2	—	0.6	0.8
30	0.1		0.05	0.0	0.1	0.2	0.2	—	0.2	0.1
35	0.1		0.0	0.0	0.1	0.1	0.2	—	0.1	—

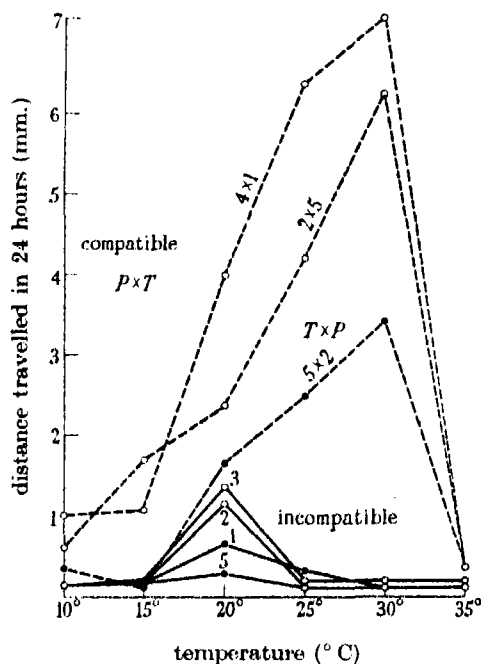


FIGURE 6

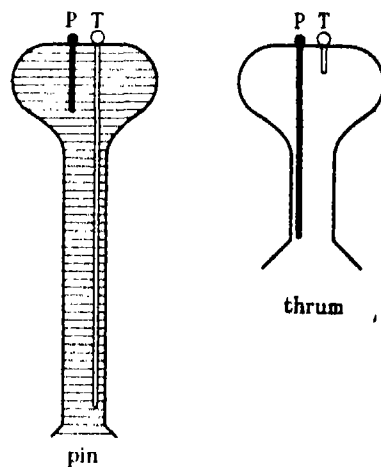


FIGURE 7

FIGURE 6. Length of pollen tubes in *Primula obconica*, 24 hr. after pollination, showing the different effect of temperature on legitimate (compatible) and illegitimate (incompatible) tubes. A difference in growth between reciprocal pollinations is also shown. Thrum pollen grows faster than pin pollen in legitimate pollinations, slower in illegitimate.

legitimate	illegitimate
---○--- pin x thrum	—○— pin selfed
---●--- thrum x pin	—●— thrum selfed

FIGURE 7. Pollen-tube growth in pin and thrum styles in *Primula obconica* at the optimum temperature for growth. Pin pollen grows slower than thrum in legitimate pollinations but faster in illegitimate.

temperature at the time of pollen-tube growth. In this point too there is agreement with *Oenothera*.

Two points from the results in *Primula obconica*, although not directly connected with the main object of the experiment, are worth noting: (1) Despite the fact that the pin style is twice as long as the thrum style, legitimate pollen tubes take approximately the same time to reach the ovary. In other words, thrum pollen grows twice as fast as pin in legitimate pollinations. (2) Thrum pollen grows at nearly half the speed of pin pollen in illegitimate pollinations (figure 7).

TABLE 7. POLLEN-TUBE GROWTH OF SELF-POLLINATIONS IN *PRIMULA OBCONICA* AFTER PRETREATMENTS OF 48 HR. AT 20 AND 30° C, SHOWING THAT THE POLLEN TUBES GROW NEARLY AS FAST AFTER A PRETREATMENT OF 30° C AS AFTER ONE OF 20° C (MM.)

pretreatment (°C)	20	20	30
after pollination (°C)	30	20	20
plant 3K	0.1	0.5	0.4
	0.2	0.7	0.9
		0.7	
plant 2K	—	0.9	0.4
		1.5	
plant 4K	0.2	2.5	2.1
		3.0	

The fact that thrum pollen is more inhibited than pin pollen in illegitimate matings, despite the more vigorous growth of thrum pollen, is probably brought about by natural selection. As the thrum style is only half as long as the pin, the thrum pollen must be twice as strongly inhibited to give the same security against self-fertilization.

6. *PRIMULA SINENSIS*

Primula sinensis is heterostyled in a similar way to *P. obconica*, but illegitimate matings, though less fertile than legitimate, are not completely sterile. This degree of illegitimate sterility seems to be an adaptation to the optimum amount of out-breeding, and it has been possible to change the degree of sterility by selection (Mather & de Winton 1941). It is of interest therefore to determine whether this reduced fertility is due to an inhibition of pollen-tube growth similar to but weaker than that in *P. obconica*.

The effect of temperature on pollen-tube growth in legitimate matings is strikingly different in the two crosses (figure 8). In pin × thrum, the optimum temperature is not clearly defined; in one pollination it is 20° C, in another 25° C and in a third 30° C. Also there is not a rapid fall in the growth rate at temperatures above the optimum. In thrum × pin there is a very definite optimum temperature at about 20° C. These differences are remarkable; the pin × thrum case is somewhat similar to compatible pollination curves in other plants except that the

increased rate of growth at temperatures above 20° C seems to be offset by intolerance of the plant to high temperatures. The thrum $\times$ pin curve is quite unlike any compatible curve in other plants and is more like the curve for incompatible tubes.

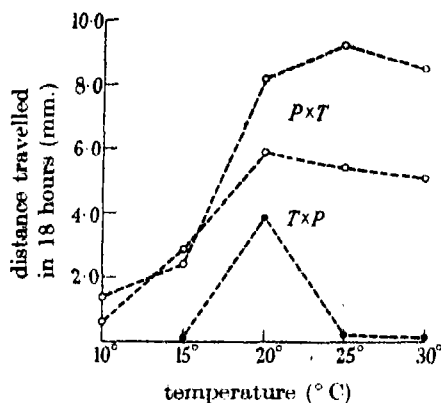


FIGURE 8

FIGURE 8. Length of legitimate pollen tubes 18 hr. after pollination in *Primula sinensis*.
 ---○--- pin $\times$ thrum, ---●--- thrum $\times$ pin.

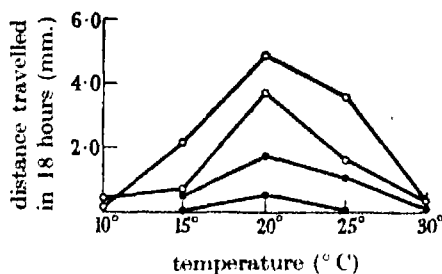


FIGURE 9

FIGURE 9. Length of illegitimate pollen tubes 18 hr. after pollination in *Primula sinensis*.
 —○— pin selfed, —●— thrum selfed.

The effect of temperature on pollen-tube growth in illegitimate matings is very similar to that in *P. obconica*, the optimum temperature being 20° C for all pollinations (figure 9). Comparing the mean pollen-tube growth of all legitimate and illegitimate pollinations at optimum temperatures after 18 hr., there is considerable difference as shown below:

illegitimate		legitimate	
pin selfed	4.4 mm.	pin $\times$ thrum	8.4 mm.
thrum selfed	0.8 mm.	thrum $\times$ pin	3.2 mm.

The thrum pollen grows faster than pin pollen in legitimate matings, pin faster than thrum in illegitimate matings, as in *P. obconica*. However, if the growth rate of pin pollen on pin style is compared with that on a thrum style, it is found that the pin pollen actually grows faster on its own type style. Thus it would appear that there is no physiological inhibitory action between pin pollen and pin style. But there is an inhibition of thrum pollen on a thrum style. This gives some information on the mode of evolution of the physiological bar to fertilization in illegitimate matings. Pin pollen is less vigorous than thrum, as shown by comparisons of the two legitimate pollinations. The pin pollen is adapted for growing down a short style; consequently even without a physiological inhibitory mechanism it is not efficient as a pollinator in its own style, which is long. This inefficiency alone affords some degree of incompatibility, as in *P. sinensis*.

The thrum pollen, on the other hand, being adapted to grow down a long style, must be physiologically inhibited to give even the slightest degree of infertility on its own style. Any further development towards complete sterility of illegitimate matings will require the inhibitory mechanism in pin $\times$ pin matings and a strengthening of this mechanism in thrum $\times$ thrum matings, as in *P. obconica*.

The adaptation for growing down a long style is brought about by the larger size of thrum pollen in most heterostyled plants. In the heterostyled *Linum grandiflorum*, which has pollen of one size, the adaptation, I find, is brought about by the higher osmotic pressure of the thrum pollen.

7. THE NATURE OF THE INCOMPATIBILITY REACTION

The extremely specific nature of self-incompatibility in plants has led to the conclusion that the inhibition of pollen tubes through incompatibility is due to a reaction of the immunity type (East 1929; Sears 1937). If this view is correct, incompatibility is due to the immunity reaction of specific proteins or polysaccharides. This is, indeed, the only plausible explanation in view of the large number of different incompatibility allelomorphs present in some plants (45 in *Oenothera organensis* (Emerson 1940), cf. *Trifolium pratense* (Williams 1939)). The effect of temperature on pollen-tube growth gives further evidence. The rate of incompatible-tube growth is slowed down by high temperatures. This can be interpreted as evidence for increased precipitation of the specific proteins at higher temperatures. This is known to be the case in animal sera (cf. Marrack 1938).

Furthermore, in no plants tested is there evidence for a particular 'zone' in the style in which the reaction takes place. The reaction is a mutual one between the tube and all the tissue of the style in the plants tested.

Plants with self-incompatibility of the S type demand great variation and specificity of reaction, whereas distylic plants such as *Primula* species need only two different reactions. Such limited variation could be met by osmotic pressure or pH differences in the styles and pollen grains. Therefore it might be expected that all self-incompatible species of the S system will have an immunity reaction, while in distylic plants other mechanisms may be found as well. However, in the case of *Primula obconica* and *P. sinensis* there seems to be little doubt that the mechanism is of the immunity type, since the temperature effect is the same as in *Oenothera* and *Prunus*. In *Linum grandiflorum*, on the other hand, which is likewise distylic, no effect of temperature could be found on the rate of incompatible pollen-tube growth, and the mechanism is probably one of osmotic pressure differences. However, a lack of response to temperature is not an infallible criterion of the type of reaction, because, as in certain genotypes of *Oenothera*, this may be so strong even at low temperatures that its further increase will not effect pollen-tube growth.

REFERENCES

- Buchholz, J. T. & Blakeslee, A. F. 1927 Pollen-tube growth at various temperatures. *Amer. J. Bot.* **14**, 358-369.
- Crane, M. B. & Brown, A. G. 1937 Incompatibility and sterility in the sweet cherry, *Prunus avium* L. *J. Pomol. Hort.* **15**, 86-116.
- Crane, M. B. & Lawrence, W. J. C. 1931 Sterility and incompatibility in diploid and polyploid fruits. *J. Genet.* **24**, 97-107.
- East, E. M. 1923 Genetical aspects of self- and cross-sterility. *Amer. J. Bot.* **10**, 468-473.
- East, E. M. 1929 Self-sterility. *Bibliogr. genet.* **5**, 331-370.
- East, E. M. 1934 Norms of pollen-tube growth in incompatible matings of self-sterile plants. *Proc. Nat. Acad. Sci., Wash.*, **20**, 225-230.
- East, E. M. & Mangelsdorf, A. J. 1925 A new interpretation of the hereditary behaviour of self-sterile plants. *Proc. Nat. Acad. Sci., Wash.*, **11**, 166-171.
- Emerson, S. H. 1938 The genetics of self-incompatibility in *Oenothera organensis*. *Genetics*, **23**, 190-202.
- Emerson, S. H. 1940 Growth of incompatible pollen tubes in *Oenothera organensis*. *Bot. Gaz.* **101**, 890-911.
- Kakizaki, Y. 1930 Studies on the genetics and physiology of self- and cross-incompatibility in the common cabbage (*Brassica oleracea* L. var. *capitata*). *Jap. J. Bot.* **5**, 133-208.
- Lawrence, W. J. C. 1930 Incompatibility in polyploids. *Genetica*, **12**, 269-296.
- Marrack, J. R. 1938 The chemistry of antigens and antibodies. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, no. 230 (H.M.S.O.).
- Mather, K. & de Winton, D. 1941 Adaptation and counter-adaptation of the breeding system in *Primula*. *Ann. Bot., Lond.*, **5**, 297-311.
- Roy, B. 1939 Studies on pollen tube growth in *Prunus*. *J. Pomol. Hort.* **16**, 320-328.
- Sears, E. R. 1937 Cytological phenomena connected with self-sterility in the flowering plants. *Genetics*, **22**, 130-181.
- Williams, R. D. 1939 Incompatibility alleles in *Trifolium pratense* L.; their frequency and linkage relationships. *Proc. 7th Int. Genet. Congr. Edinb.*
- Yasuda, S. 1930 Physiological researches. VIII. On the self-fertilising ability of flowers in the bud of the self-incompatible plants. *Bot. Mag., Tokyo*, **44**, 678-687.
- Yasuda, S. 1934 Physiological researches on self-incompatibility in *Petunia violacea*. *Bull. Imp. Coll. Agric. For. Morioka*, **20**, 1-95.

Some experiments on the trichromatic theory of vision

By H. V. WALTERS

Imperial College of Science and Technology, Technical Optics Section

(Communicated by H. Hartridge, F.R.S.—Received 26 February 1942)

A repetition of an investigation carried out by Wright in 1934 is made with an improved design of the Wright trichromatic colorimeter. The technique is described and the assumptions in the method for determining the fundamental response curves discussed. New observations are reported in which the generalized form of von Kries's law of coefficients is found to break down, and the departures from the coefficient law are analysed. Possible physiological causes for the breakdown are discussed. By a modification in the calculations, a new attempt has been made to locate the colours in the trichromatic colour chart corresponding to the three fundamental responses, and to derive the response curves.

Calculations were first made in the W. D. Wright trichromatic system and were transformed into the rectangular uniform chromaticity scale system of Breckenridge & Schaub, showing the caution required in interpreting the statistical spread of the results.

HISTORICAL INTRODUCTION

A number of investigators, adopting the Young-Helmholtz theory of normal trichromatic colour vision, have attempted to locate the three postulated fundamental sensations in the colour diagram, with their corresponding response curves. The methods employed have included:

A. Speculation from the form of the spectral locus in the colour triangle (Clerk Maxwell 1890).

B. A combination of A with additional data on discrimination (Helmholtz 1911, Hecht 1932).

C. A combination of A with additional experimental work on luminosities of mixtures of spectral lights (Abney 1913).

D. Consideration of the colour characteristics of trichromats and dichromats, regarding the latter group as having one or other of the fundamental sensations missing (König 1929).

E. Measurement of changes in the appearance of stimuli under adaptation to strongly chromatic light, by a binocular matching method (Wright 1934).

F. The use of the Stiles & Crawford effect (directional sensitivity of the retina) (Stiles 1939).

G. Analysis of electroretinograms of mammals, amphibians and fishes using a microelectrode technique (Granit 1941).

The various sets of response curves which are derived by the above methods show marked differences, though some agree tolerably well, and it was felt that the uncertainty existing about the sensations postulated by the theory might be due

to inadequate experimental data, quite apart from the possibility of a faulty theory (see figure 9).

The adaptation method (E above) was used by Wright with positive, though not entirely satisfactory results. A description of these experiments formed part of a paper on colour adaptation (Wright 1934). A considerable part of this paper was devoted to the establishment of an extended form of von Kries's 'Law of Coefficients' (von Kries 1904), the truth of which generalization (Wright's law) is fundamental to this method for deriving the sensations. The main purpose of this paper is further to discuss the coefficient law, which is stated in symbolic form as assumption (3) in the discussion of the method for determining the fundamental responses (p. 31).

APPARATUS

The apparatus provides a small illuminated field for each of the observer's eyes, one of which can be varied in colour and intensity by the operator (the test patch), and the other by the observer (the comparison patch). This latter is adjusted until a complete binocular match is obtained. The fields are illuminated by narrow wave-length bands of the spectrum of a tungsten ribbon-filament lamp, the test patch usually being one such band selected in the region of a known wave-length, while the comparison patch is a mixture of three wave-length bands, usually chosen to lie in the red, green and blue region of the spectrum at 0.65 , 0.53 and 0.46μ respectively. The intensities of each of these contributions to the colour of the comparison patch is independently controlled by the observer.

In addition, it is also necessary for the operator to be able to cut off the light from the test and comparison patches, and substitute in the right eye (otherwise receiving the test patch) a large circular field of relatively high intensity and strongly saturated light. This is called the adaptation patch.

The apparatus meeting these requirements is described in the *Journal of Scientific Instruments* (Wright 1939), which should be consulted for details. This trichromatic colorimeter provides rectangular test and comparison patches of angular size $2 \times 1^\circ$, and for a person with normal muscle balance the comparison patch appears vertically over the test patch and separated from it by approximately 1° visual angle. The intensity controls for the red, green and blue components of the comparison patch are situated so that their adjustment involves a minimum of fatigue. The observer's head is held steady by biting on to a wax dental impression attached to the instrument, and adjustments are provided so that the exit pupils of the instrument may be centred over the eye pupils, using the chromatic aberration of the eye as a test for centring. When centring is defective, the fields are unsymmetrically fringed with colour, the fringes disappearing when centring is complete. In this way troubles due to the 'Stiles and Crawford effect' (1933) are eliminated.

The adaptation patch is of about 6° diameter, its colour and intensity being varied by suitable combinations of Ilford 'Spectrum' and 'Neutral Density' filters.

The change-over from viewing the test and comparison patches to viewing the adaptation patch is effected by a moving mirror and shutter system controlled by the operator. Figure 1*a* indicates the relative positions of the test, comparison, and adaptation patches.

Variation in horizontal muscle balance was quite common and led to displacements such as those shown in figure 1*b*, and after some time vertical displacements began to be troublesome on some occasions, the patches slowly moving together and finally fusing (figure 1*c*), when matching was, of course, impossible. When this occurred the experimental run was abandoned.

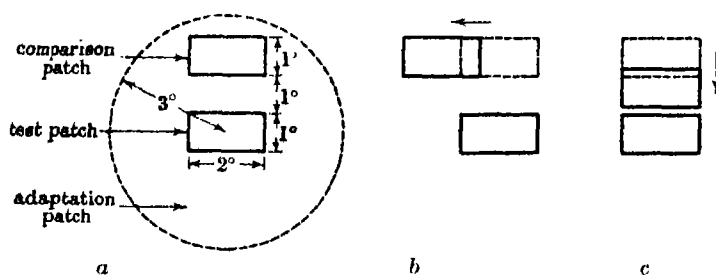


FIGURE 1. Appearance of binocular matching fields. (a) Relative position of test, comparison, and adaptation patches. (b) Effect of horizontal muscle imbalance. (c) Effect of vertical muscle imbalance.

BINOCULAR MATCHING TECHNIQUE

Before discussing the theory of the method for determining the fundamental responses, we shall give an account of one type of experiment in summarized form, subsequently explaining the reasons for each step.

- (1) The observer adjusts the instrument to suit his interocular distance, etc.
- (2) Both eyes are dark adapted for 20 min.
- (3) Binocular matches using 0.494 and 0.5825 μ as test patches are made at fairly low intensities (approx. 20 photons).
- (4) Binocular matches are made of all the test stimuli to be investigated, the physical intensities being reduced by calibrated neutral filters or sectors. These are known as the 'initial' or 'pre-adaptation' matches.
- (5) To facilitate subsequent matching the comparison patch is adjusted to the probable appearance of a given test stimulus after adaptation.
- (6) The right eye is light adapted with moderately intense saturated light for 3 min., this being sufficient to render it in a steady state of adaptation as gauged by binocular matching methods (Wright 1934).
- (7) After 3 min. the light adaptation is interrupted by the operator, who, using a stop-watch, now shows the observer the test and comparison patches for 1 sec. in every 10, until the latter has obtained a satisfactory match. A second match is then made after slightly upsetting the first, and the mean readings of the two matches taken.

(8) The stop-watch hands are returned to zero, then simultaneously the adaptation beam is cut off permanently and the watch started. The test and comparison patches are now continuously visible.

(9) The observer makes a series of rapid matches as the sensitivity of the right eye recovers, all the while taking care not to look at the patches for longer than necessary to make matches in comfort, experiment showing this to yield most consistent results, without appreciable systematic error.

After a short period of rest, another observational run is made (operations (5)–(9)), say for a different test stimulus or a different intensity of the same test stimulus.

Whenever several test stimuli of different intensities are being investigated during one observing session, every run is repeated if possible, and the repeated runs arranged in reverse order, so that if any continuous and fairly uniform change takes place in the physiological state of the retina (e.g. dark adaptation of the rods), its effects are minimized by averaging.

It is found experimentally that after exposure to normal intensities of artificial light 20 min. dark adaptation is sufficient to bring the eye to a steady state of sensitivity, especially as we are concerned with cone vision (Wright 1934). Thus process (2) ensures that all the 'initial' or 'pre-adaptation' matches are made under constant sensitivity.

The third process is necessary to fix the trichromatic units in which the results are measured. The W. D. Wright (W.D.W.) system has been defined so that there are equal numbers of units of the red and green instrument primaries (0.65 and 0.53μ) used to match 0.5825μ , and equal numbers of green and blue (0.53 and 0.46μ) to match 0.4940μ . This system is more suitable for physiological investigations than the more usual ones having white as a basic stimulus, the reasons being set out in a previous paper (Wright 1929).

Low intensities of the various test colours are used mainly to obviate as far as possible any light adaptation of either eye, and also because the subjective brightness in the post-adaptation matches will be low, and this accords with good physical practice.

Stage (5) is included to save time, since adjustment is necessarily more tedious during process (7). Interruption of light adaptation for 1 sec. in every 10 produces no significant change in retinal sensitivity over a period of 3 min. or so, but the interruption period may not be safely lengthened, while shortening produces deterioration in the observations.

We shall see that an assumption which is basic to the method for determining the fundamental responses is that the eye can be maintained very nearly in a constant state of adaptation to one field while another dissimilar field is presented to it sufficiently long for an estimate of its intensity and colour to be made. That this assumption is an ideal only approximated to is an inherent consequence of the fact that the eye does adapt itself to different fields. The effect of the deviation from ideal would not matter if the state of adaptation varied in a constant manner

when the test stimulus is substituted for the adapting stimulus, i.e. independently of the nature of the test stimulus. Judging from our knowledge of the rates of the light- and dark-adaptation processes, this will be most nearly true when the test stimulus is of much lower intensity than the adapting stimulus, and of course when the variation between the several test patches is not large.

It has been found (Schouten & Ornstein 1939) that very rapid (α -) adaptation takes place on presentation to the eye of a new stimulus, and that thereafter a relatively slow (β -) adaptation sets in. We must anticipate that α -adaptation may be highly specific, that is, the colour as well as the intensity of the stimulus playing a major role in determining the state of sensitivity of the retina. It is too much to hope that with the present technique it will be possible to eliminate or even minimize this factor experimentally. If the presentation time for the test stimulus is reduced, the advantage to be expected in respect of α -adaptation is more than offset by the poor discrimination entailed. Its possible presence, therefore, has to be considered in analysing the results.

MATHEMATICAL TREATMENT OF THE METHOD FOR DETERMINING THE FUNDAMENTAL RESPONSES

The following three assumptions were made in the original determinations made by Wright:

(1) The state of sensitivity of the right eye under light adaptation of about 2000 photons is not sensibly altered by the presentation for 1 sec. in every 10 of a different stimulus of somewhat lower intensity; say up to 500 photons.

(2) The left eye is not appreciably reduced in sensitivity, and in particular does not undergo colour adaptation, upon presentation of the comparison patch, which is frequently highly saturated.

(3) Suppose the two eyes to be maintained in constant states of adaptation, either of which may lie within a range from complete dark adaptation to that of light adaptation with monochromatic red or green light of the order of 10,000 photons. Then if each of two pairs of stimuli provides equality of sensation (a 'binocular match'), the resultant pair obtained by physical addition of the other two pairs will itself produce a binocular match. In symbols, if we represent a stimulus by Σ , differentiating between those stimuli presented to the left and right eyes by suffixes l and r , and between different stimuli to the same eye by the prefixes 1, 2, etc., then if

$$[{}_1\Sigma_l]_A = [{}_1\Sigma_r]_B \quad \text{and} \quad [{}_2\Sigma_l]_A = [{}_2\Sigma_r]_B,$$

it follows that

$$[{}_1\Sigma_l + {}_2\Sigma_l]_A = [{}_1\Sigma_r + {}_2\Sigma_r]_B.$$

The sign of equality signifies a binocular match, and the square brackets with suffixes A and B indicate that the left and right eyes are in states of adaptation A and B respectively.

This condition is a generalization requiring the fulfilment of conditions (1) and (2), and may be described as the 'generalized coefficient law'.

We shall now see how these assumptions enter the arguments involved in the method, as we develop it.

There is considerable circumstantial evidence for the view that wave-lengths between approximately 0.54μ (slightly yellow-green) and the extreme red end of the spectrum effectively elicit only the red and green responses. For the sake of convenience in using the W.D.W. trichromatic system having 0.53μ as one instrument primary, the range may be extended to that wave-length without serious error. This evidence consists of the linearity of the spectral locus over the range indicated, and the fact that blue adaptation does not make, say a yellow at 0.58μ appear too saturated to be binocularly matched by a mixture of the W.D.W. red and green primaries (0.65 and 0.53μ). Both these arguments are reasonable, although not conclusive. Taking them on trust, we can simplify the search for the red and green responses.

Let us represent units of red, green and blue instrument primary by R , G and B , and the number of each in a match as α , β and γ respectively.

Suppose then that we represent binocular matches on two test stimuli in the red-yellow-green region, in quality and intensity by the equations

$$[{}_1C_r]_D = [\alpha_1 R_l + \beta_1 G_l]_D, \quad (i)$$

$$[{}_2C_r]_D = [\alpha_2 R_l + \beta_2 G_l]_D, \quad (ii)$$

where $[]_D$ indicates a state of dark adaptation. After adaptation, let the matches be represented by

$$[{}_1C_r]_A = [\alpha'_1 R_l + \beta'_1 G_l]_D, \quad (iii)$$

$$[{}_2C_r]_A = [\alpha'_2 R_l + \beta'_2 G_l]_D. \quad (iv)$$

It should be noted that though the equations are equations of sensation, we believe this to involve equality of response, and we can proceed without any assumption about, or reference to, the response-sensation mechanism.

Suppose we now consider a hypothetical test stimulus represented by

$$[P_r]_D = [xR_l + yG_l]_D, \quad (v)$$

which also satisfies the condition that

$$[P_r]_A = k[xR_l + yG_l]_D, \quad (vi)$$

k being a numerical constant. If we could obtain a stimulus evoking only one of the fundamental responses, it would obviously satisfy this condition, and our task is essentially the location of such a stimulus. This statement is the keystone of the argument, and in it lies the distinctiveness of the method. We can see without rigorous proof that should the adaptation be such that some real stimulus satisfied the condition, all stimuli would do so. The existence in practice of colour changes ($\alpha/\beta \neq \alpha'/\beta'$) under the adaptations used is a guarantee against this latter possibility.

All the previous equations represent binocular colour matches, but we can combine equations (i), (ii), and (v) to obtain an equation representing a monocular colour match, i.e.

$$[P_r]_D = [\mu_1 C_r + \nu_2 C_r]_D, \quad (\text{vii})$$

where, comparing coefficients,
$$\mu = \frac{x\beta_2 - y\alpha_2}{\alpha_1\beta_2 - \alpha_2\beta_1}, \quad (\text{viii})$$

$$\nu = \frac{y\alpha_1 - x\beta_1}{\alpha_1\beta_2 - \alpha_2\beta_1}. \quad (\text{ix})$$

Utilizing a property of the monocular colour match

$$[P_r]_A = [\mu_1 C_r + \nu_2 C_r]_A \quad (\text{x})$$

$$= k[\mu_1 C_r + \nu_2 C_r]_D. \quad (\text{xi})$$

Combining equations (i), (ii), (iii), (iv), (x) and (xi) we obtain the following equation uniquely determining the colour of the fundamental responses:

$$\frac{x}{y} = \frac{x(\alpha'_1\beta_2 - \alpha'_2\beta_1) + y(\alpha_1\alpha'_2 - \alpha'_1\alpha_2)}{x(\beta'_1\beta_2 - \beta_1\beta'_2) + y(\alpha_1\beta'_2 - \alpha'_2\beta_1)}. \quad (\text{xii})$$

For the purpose of location in the colour diagram, we are only concerned with the ratio of x to y , and we have in (xii) a quadratic equation in that ratio, the two roots of which clearly should give us the fundamental red and green responses, H_r and J_r , of the right eye.

Writing
$$H_r = x_1 R_1 + y_1 G_1, \quad J_r = x_2 R_1 + y_2 G_1, \quad (\text{xiii})$$

we can find x_1/y_1 , x_2/y_2 .

Recasting (xii) and solving:

$$\frac{x}{y} = \frac{(\alpha_2\beta'_1 - \alpha_1\beta'_2 + \alpha'_1\beta_2 - \alpha'_2\beta_1) \pm \sqrt{(\alpha_2\beta'_1 - \alpha_1\beta'_2 + \alpha'_1\beta_2 - \alpha'_2\beta_1)^2 - 4(\beta'_1\beta_2 - \beta_1\beta'_2)(\alpha'_1\alpha_2 - \alpha_1\alpha'_2)}}{2(\beta'_1\beta_2 - \beta_1\beta'_2)}. \quad (\text{xiv})$$

This rather unwieldy equation is considerably simplified when either one of the α 's or one of the β 's is zero. These important cases arise when, with red adaptation, one of the test stimuli is 0.65μ , and with green adaptation, 0.53μ (0.53μ may not be used as a test stimulus with red adaptation, nor 0.65μ with green adaptation, since neither can then be matched binocularly, appearing more saturated than any possible mixture of spectral primaries in the left eye).

The two special forms are then

$$\frac{x}{y} = \frac{(\alpha_1\beta'_2 + \alpha'_2\beta_1 - \alpha_2\beta'_1) \pm \sqrt{(\alpha_1\beta'_2 + \alpha'_2\beta_1 - \alpha_2\beta'_1)^2 + 4\beta_1\beta'_2(\alpha'_1\alpha_2 - \alpha_1\alpha'_2)}}{2\beta_1\beta'_2}. \quad (\text{xiv a})$$

in the case of red adaptation ($\beta_2 = 0$), and

$$\frac{y}{x} = \frac{(\alpha_1\beta'_2 + \alpha'_2\beta_1 - \alpha'_1\beta_2) \pm \sqrt{(\alpha_1\beta'_2 + \alpha'_2\beta_1 - \alpha'_1\beta_2)^2 + 4\alpha_1\alpha'_2(\beta'_1\beta_2 - \beta_1\beta'_2)}}{2\alpha_1\alpha'_2}. \quad (\text{xiv b})$$

for green adaptation ($\alpha_2 = 0$). It is simpler to evaluate y/x in this latter case.

Since the expression for x/y is homogeneous and symmetrical in the measured quantities, there is no point in aiming at a selectively high standard of accuracy in any of the measurements. The accuracy of the value of, say, α'_1 varies inversely as its ratio to β'_1 , so it is desirable not to have any of the quantities very small. The possible range of test colours for red adaptation is roughly from 0.65 to 0.5825μ , and this is shortened in practice to 0.61μ , while the useful range for green adaptations was found to be from 0.53 to 0.5825μ , the latter being a useful test stimulus in reducing by one the number of initial matches necessary.

Before proceeding to discuss the results actually obtained, we can investigate the effects to be expected from particular errors in the measured quantities.

We may say straight away that we expect the red and green fundamentals, when expressed in terms of spectral primaries (0.65 and 0.53μ) to have negative green and red coefficients respectively, as a natural consequence of our basic theory, in which the responses are all positive. Hence we should expect both values of x/y to be real and negative, one being numerically greater than unity, and the other less.

Condensing (xiv), (xiv*a*), (xiv*b*) into the usual notation

$$\frac{x}{y} \left(\text{or } \frac{y}{x} \right) = - \frac{b \pm \sqrt{(b^2 - 4ac)}}{2a},$$

where a , b and c can easily be recognized, we see that a , b and c should all be of the same sign and $b^2 > 4ac$. In the special cases we have made the arbitrary choice that a , b and c should all be positive, by suitably associating suffixes with test stimuli.

The following conditions are therefore necessary to satisfy the logical demands of the trichromatic theory:

$$\alpha_1\beta'_2 + \alpha'_2\beta_1 \sim \alpha_2\beta'_1 + \alpha'_1\beta_2, \quad (\text{xv})$$

$$\beta'_1\beta_2 \sim \beta_1\beta'_2, \quad (\text{xvi})$$

$$\alpha'_1\alpha_2 \sim \alpha_1\alpha'_2, \quad (\text{xvii})$$

$$(\alpha_1\beta'_2 - \alpha_2\beta'_1 + \alpha'_2\beta_1 - \alpha'_1\beta_2)^2 > 4(\beta_2\beta'_1 - \beta_1\beta'_2)(\alpha_1\alpha'_2 - \alpha'_1\alpha_2), \quad (\text{xviii})$$

the signs of the inequalities in (xv), (xvi) and (xvii) being the same.

We should obviously try to make the terms on left- and right-hand sides differ by as large an amount as possible relative to their own magnitudes. For no colour change, i.e. no light adaptation, the differences vanish, and x/y becomes indeterminate, so that, *ceteris paribus*, we should aim at maximum colour change. The fulfilment of this requirement is compromised by a number of disadvantages attached to the conditions for great colour change. At very high intensities of adaptation, anomalous effects occur, including monocular colour-match breakdown, and the disturbing appearance of after-images, apart from the discomfort of the observer and the poor discrimination when the test stimulus is of much lower intensity than the adaptation. We must certainly not work too near the threshold of sensation, since the mere existence of a finite threshold seriously limits the validity of the assumptions.

RESULTS

Having acquired sufficient skill in binocular matching, a number of preliminary experiments were made using both red and green adaptations and the results reviewed. Quite apart from the large statistical spread, it was clear that the values of H and J determined from red and green adaptations did not agree. In fact, most of the red adaptations yielded complex values of the x/y ratio, indicating some serious systematic error.

These experiments were made with adaptation intensities of 1000 photons for the red and 600 photons for the green, the test stimuli intensities being of the order of 200 photons. The adaptation intensities were doubled, but the results were not informative. All the results from red adaptation were now found to be complex.

The original determinations (Wright 1934), giving real and reasonable results, had been made with a somewhat different technique, in which it was only possible to obtain points on the recovery curves after some 20 sec. The complete recovery curves which it was now possible to draw with the new technique were investigated to see if they showed some systematic error, as gauged by extrapolation, at $t = 0$. The positive blue, for instance, is relatively large at $t = 0$, but under all conditions is found to disappear very quickly and to be quite negligible for our present purposes at $t = 10$ sec., so that its presence would not be suspected from extrapolation. No definite error could be observed, and, further, values of H and J obtained from α'_n, β'_n , read from the recovery curves at 10 and 20 sec., showed no improvement in agreement or significant change in location. A representative set of results so far described is set out below (table 1).

The fact that Wright had obtained reasonable, and the author unreasonable, results could be explained by differences in dynamic characteristics of vision, giving fortuitously good results in the former's experiments. These differences were known to exist.

It was felt that the positive blue phenomenon which has been mentioned might be of considerable significance in connexion with the anomalous results. Whenever a test colour in the red-yellow-green range is viewed immediately after a strong red or green adaptation, it appears quite desaturated or maybe purplish, indicating that the blue response is being fairly strongly evoked. During recovery the amount of blue required in the binocular match rapidly subsides and disappears. An important point is also that for the same state of adaptation the positive blue is less in relative value the higher the intensity of the test stimulus.

As a working hypothesis, suppose that the positive blue is a superposition on a response obeying the coefficient law, as an 'after-image' or other allied phenomenon. If the blue response alone shows this anomaly, the effect on the measurements with red and green instrument primaries due to the non-identity of the blue instrument primary with the blue fundamental response would almost certainly be too small to account for the results described.

TABLE 1

recovery time sec.	adaptation colour and intensity	pair of test stimuli with intensities in photons		z_1/y_1	z_2/y_2	H (if real) (W.D.W. system)	J (if real) (W.D.W. system)	H (R.U.C.S. system)	J (R.U.C.S. system)
0	red, 1000 photons	0.5825 μ (400)	0.65 μ (300)	-6.23 $\pm$ 0.31i	-	-	-	-	-
		0.61 μ (450)	0.65 μ (300)	-4.81 $\pm$ 1.8i	-	-	-	-	-
10		0.5825 μ (400)	0.65 μ (300)	-4.72 $\pm$ 0.11	1.27R -0.27G	0.1R +0.9G	0.08U -0.77V +1.69W	0.07U +0.12V +0.81W	-
		0.61 μ (450)	0.65 μ (300)	-7.36 $\pm$ 0.29i	-	-	-	-	-
0	red, 2000 photons	0.5825 μ (400)	0.65 μ (300)	-8.3 $\pm$ 0.89i	-	-	-	-	-
		0.61 μ (450)	0.65 μ (300)	-5.9 $\pm$ 1.42i	-	-	-	-	-
10		0.5825 μ (400)	0.65 μ (300)	-6.2 $\pm$ 4.6i	-	-	-	-	-
		0.61 μ (450)	0.65 μ (300)	-9.5 $\pm$ 5.8i	-	-	-	-	-
0	green, 600 photons	0.53 μ (220)	0.5825 μ (400)	-3.9	-0.42	1.35R -0.35G	-0.64R +1.64G	0.08U -0.87V +1.79W	0.07U +0.41V +0.52W
		0.53 μ (220)	0.56 μ (330)	-10.8	+0.16	1.10R -0.10G	0.14R +0.86G	0.08U -0.59V +1.51W	0.07U +0.10V +0.83W
10		0.53 μ (220)	0.5825 μ (400)	-5.1	-0.47	1.36R -0.36G	-0.89R +1.89G	0.08U -0.88V +1.80W	0.07U +0.50V +0.43W
		0.53 μ (220)	0.56 μ (330)	-7.9	-0.03	1.14R -0.14G	-0.03R +1.03G	0.08U -0.63V +1.55W	0.07U +0.18V +0.76W
0	green, 1200 photons	0.53 μ (220)	0.5825 μ (400)	-4.5	-0.08	1.28R -0.28G	-0.06R +1.08G	0.08U -0.78V +1.70W	0.07U +0.20V +0.73W
		0.53 μ (220)	0.56 μ (330)	-4.6 $\pm$ 0.29i	-	-	-	-	-
10		0.53 μ (220)	0.5825 μ (400)	-8.1	+0.23	1.24R -0.24G	0.19R +0.81G	0.08U -0.74V +1.66W	0.07U +0.07V +0.86W
		0.53 μ (220)	0.56 μ (330)	-6.1	-0.80	1.20R -0.20G	-0.87R +1.87G	0.08U -0.69V +1.61W	0.07U +0.51V +0.42W
		0.53 μ (140)	0.5825 μ (210)	-8.4	+0.17	1.14R -0.14G	0.15R +0.85G	0.08U -0.63V +1.55W	0.07U +0.10V +0.83W

In order to make some quantitative measurements, two separate stimuli were investigated, and then physically mixed, the 'sum' of the separate recovery curves for each component being compared with the corresponding curves for the combined stimuli. The few red adaptation experiments that were made showed that not only did the blue response behave anomalously but so also did the red and green, though the latter might not alone have aroused suspicion. Typical curves obtained in this way are shown in figure 2. Wright made a number of experiments of this type, and his results agreed in a general way with those of the author, indicating quite definitely a departure from the coefficient law. These observations, though few relative to the total number made by the author, are therefore of considerable significance as circumstantial evidence.

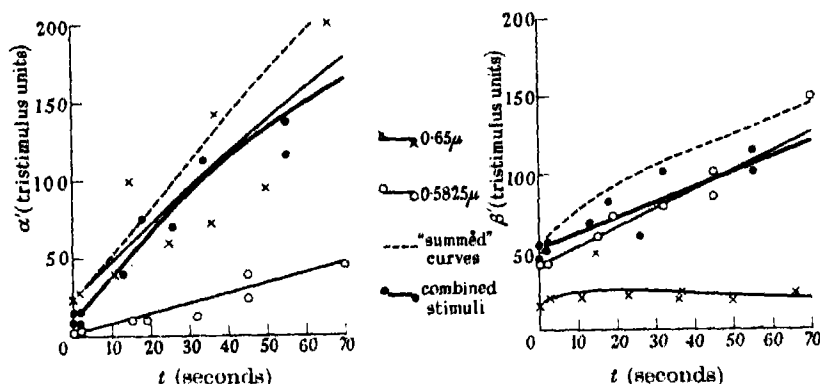


FIGURE 2. Typical recovery curves exhibiting breakdown in 'colour arithmetic'. Red adaptation 2000 photons. Test stimuli 0.5825μ , 0.65μ independently and combined.

The implication of these findings is either: (a) breakdown of the monocular colour match, or (b) breakdown of the generalized coefficient law.

The former possibility was put to a straightforward test. It was found that the adaptation intensity required to bring about breakdown in the monocular colour match was of the very high order of 20,000 photons. As we used only 2000 photons in our experiments, we had a large margin of safety.

A direct investigation of the coefficient law was now undertaken. A number of experiments were made with constant adaptation and five intensities of a test stimulus with a range of 10:1. These were obtained by the use of sectors of semi-openings 180° (i.e. no sector), 120° , 75° , 37.5° and 18° , at S , giving a convenient spacing of the intensities. The results of these new experiments are summarized below.

Red adaptation

There are large colour changes and depression of sensitivity. Positive blue is most noticeable with red test stimuli, decreasing in relative value with the wavelength of the stimulus. For any one test colour the positive blue decreases in

relative value with increasing intensity of the test stimulus, always becoming negligible after at most 10 sec.

Recovery curves, the ordinates of which are multiplied by a factor inversely proportional to the sector opening are shown below (figure 3 *a, b*) and indicate

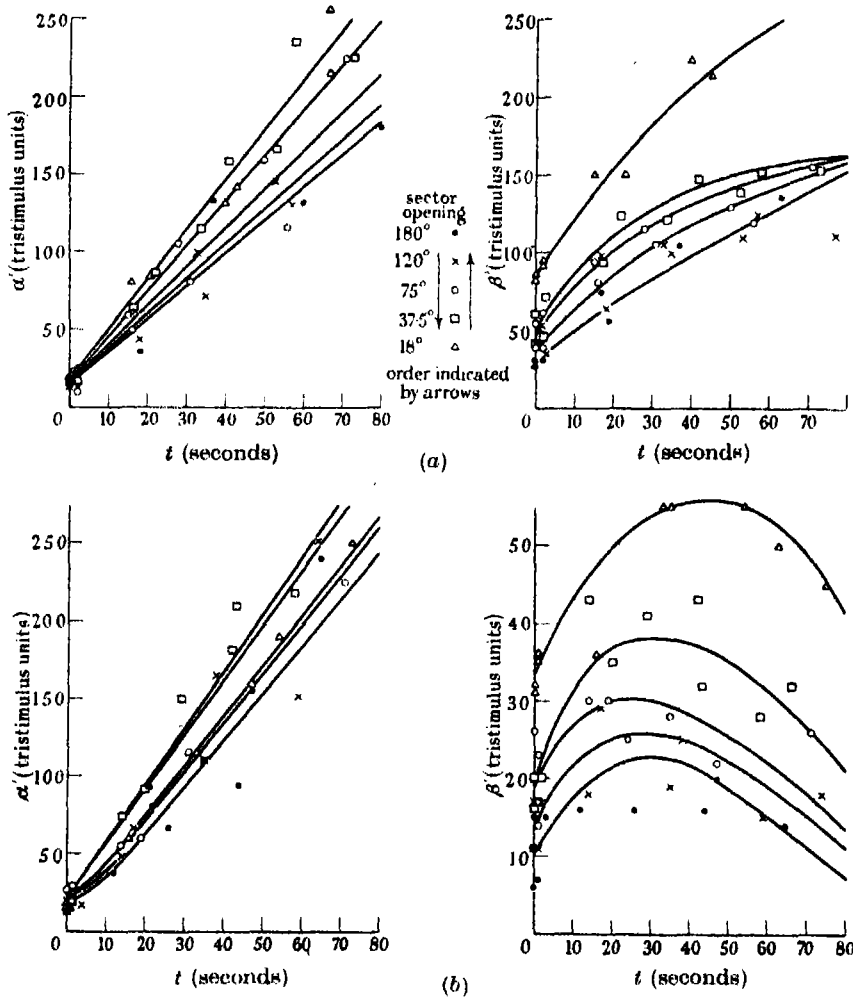


FIGURE 3. Typical recovery curves. Red adaptation 2000 photons. (a) Test stimulus 0.61μ. (b) Test stimulus 0.85μ. Ordinates multiplied by factor inversely proportional to sector angle.

systematic deviations from the coefficient law. If the law held exactly, the curves would of course all coincide. Recovery curves for the blue component are not given, since the only measurements of any significance is that at $t = 0$.

Alone, the red component recovery curves for all the test colours used appear to obey the law well within the limits of experimental error, but on taking a large number of observations, a slight deviation is exhibited, especially in test colours with considerable green in them, so that at lower intensities the values are higher

than would be given by the law. The green component curves, however, exhibit a very well-marked deviation from the law, of the same general character as the red curves show. Although Wright did not describe exactly similar experiments, he commented on apparent deviations from his law shown by the red value when a white adaptation was used. Lastly, the blue values at $t = 0$ show considerable deviations from the coefficient law.

As will be seen from the curves, the deviation of the red component from the law is not appreciable, if existent, when the test stimulus is 0.65μ , but begins to appear as the wave-length of the stimulus shortens, and at 0.5825μ is quite definite. This phenomenon is of some importance in the discussion of the results.

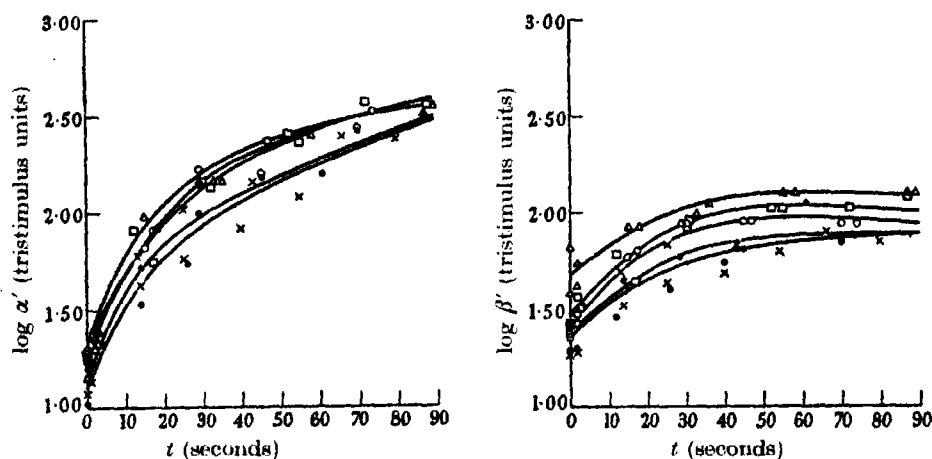


FIGURE 4. Typical recovery curves plotted logarithmically to show persistence of breakdown. Red adaptation 2000 photons. Test stimulus 0.625μ .

A further point is that many of the red recovery curves show an initial upward concavity, while none show a downward one, so that barring systematic errors, there appears to be such a departure from linearity. It must be emphasized that the deviation is of the same order as the spread of the results, and is therefore not too obvious.

The considerable amount of green in the match on 0.65μ immediately after adaptation still further increased on recovery, reaching a maximum and then gradually disappearing, being still appreciable after 100 sec. A constantly recurring phenomenon was the change in time of arriving at maximum with change of intensity, the maximum occurring later for low intensities.

It will be seen from figure 4 that the deviation from the coefficient law is quite persistent, and has only slightly decreased after 80 or 100 sec., when the sensitivity as measured by the matching stimulus has recovered to considerably more than 50 % of its dark-adapted value. Figure 4 is also typical in that it exhibits a phenomenon recurring regularly, that of the downward trend in sensitivity with successive adaptations towards the end of the experiment. This may be a fatigue effect, as distinct from adaptation.

Green adaptation

Colour changes are smaller than with red adaptation, and also the general depression of sensitivity is not so large for comparable luminosities of adapting stimulus. Positive blue is greatest in relative value with test stimulus 0.53μ , decreasing as the wave-length increases (cf. red adaptation). As with red adaptation, by the end of 10 sec. recovery time, the positive blue is insufficient to bring the subjective colour as much as a discrimination step away from the spectrum locus.

Deviations from the coefficient law for the green component are apparent for all the test stimuli used, from 0.53 to 0.5825μ , though these are not nearly as large as in the red component (figure 5).

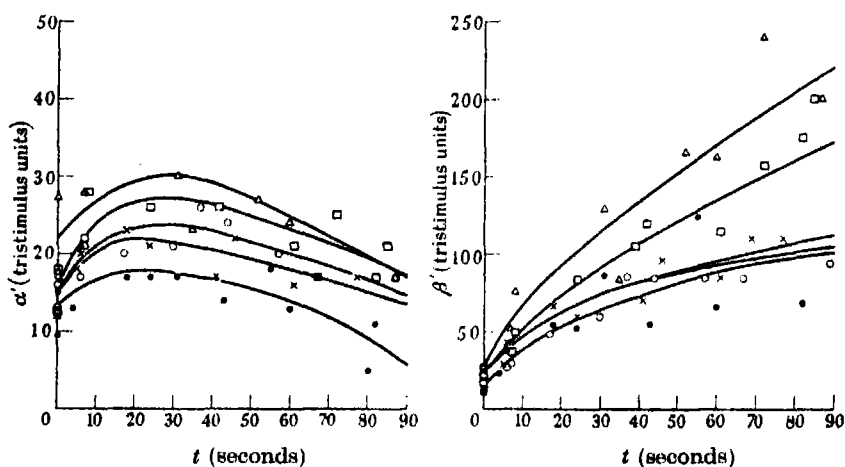


FIGURE 5. Typical recovery curves for green adaptation 1200 photons.
Test stimulus 0.53μ .

The green component shows a marked upward convexity on the early part of the curves. This could not possibly be accounted for by a fortuitous spread of the results, or systematic error.

In the red component curve for 0.53μ there is the same type of maximum as for the green in 0.65μ with red adaptation, but there is no apparent shift of the maximum with intensity. The accuracy of the curves is, however, very poor, since the red contributes very little to the luminosity, and the absence of shift should be regarded with reserve.

As with red adaptation, the deviations from the coefficient law diminish slowly with increasing recovery time, and it is known that the law holds very well when both eyes are thoroughly dark-adapted. This latter fact justifies the reduction of intensity of the test stimulus for the initial matches.

In order to gain a better picture of the way in which the matching stimulus is related to the test stimulus, smooth recovery curves are drawn, and values of the red and green components at some particular recovery time plotted against the

intensity of the test stimulus, in arbitrary units. Examples of such curves are shown in figure 6 a-c.

A single experiment was sufficient to show that with the intensities of adaptation used in these experiments, there is no 'after-image' of the adaptation when no test stimulus is presented, and the threshold of sensation is quite normal, that is to say, on raising the stimulus energy from zero there is a small uncertain range in which the stimulus becomes visible, after which the sensation increases steadily and uniformly. The curves are therefore extrapolated to pass through the origin of co-ordinates. It will be seen that the curvature increases with decreasing stimulus, so that the low intensity end is quite steeply curved.

A serious drawback in these investigations has been the fact that a considerable time is necessary to acquire facility with binocular matching, and though six other observers commenced observations, it was found impossible for them to continue, for one reason or another. No useful results were obtained, therefore, for other observers, apart from some by Wright, which checked his original determinations, and indicated the difference between his and the author's dynamic characteristics of vision.

DEDUCTIONS ABOUT THE MECHANISM

Equipped with these results, one can make some statement about the mechanism of cone vision and restrict the possibilities inherent in the Young-Helmholtz theory.

Any simple photochemical theory of vision, e.g. involving no side- or chain-reactions, leads to the generalized coefficient law as a natural consequence. In order to explain our results we must look to additional reactions of some kind, or say, to some electrical effect which may modify the responses. There is a possibility that the explanation lies in an 'after-image' of the adaptation. Direct experiment has discounted the idea of a constant after-image. In view of the facts relating to different test stimuli, we can safely rule out the possibility of an after-image varying with time, but independent of the composition and intensity of the test stimulus. It should be remarked that the term 'after-image' has been widely applied and has a very loose connotation. Here we are dealing with a complex function of the adaptation and test stimulus contributing to the response to the latter, and I shall abandon the term, anticipating our deductions, for the more suitable one of 'interaction'.

It has often been stated (Parsons 1915; Peddie 1922) after Helmholtz that the three responses are mutually independent, on the evidence of monocular colour-matching phenomena. Interaction, at any point subsequent to conversion of the stimulus energy into chemical or electrical energy, is perfectly possible within the framework of the theory, even, be it noted, if the coefficient law were perfectly true. Presumably the whole physiological states of the two retinae are very different when one is light-adapted and the other dark-adapted. It is not difficult to see how the deviations from the coefficient law become apparent under such conditions, though they cannot be detected under the usual conditions of similar adaptation.

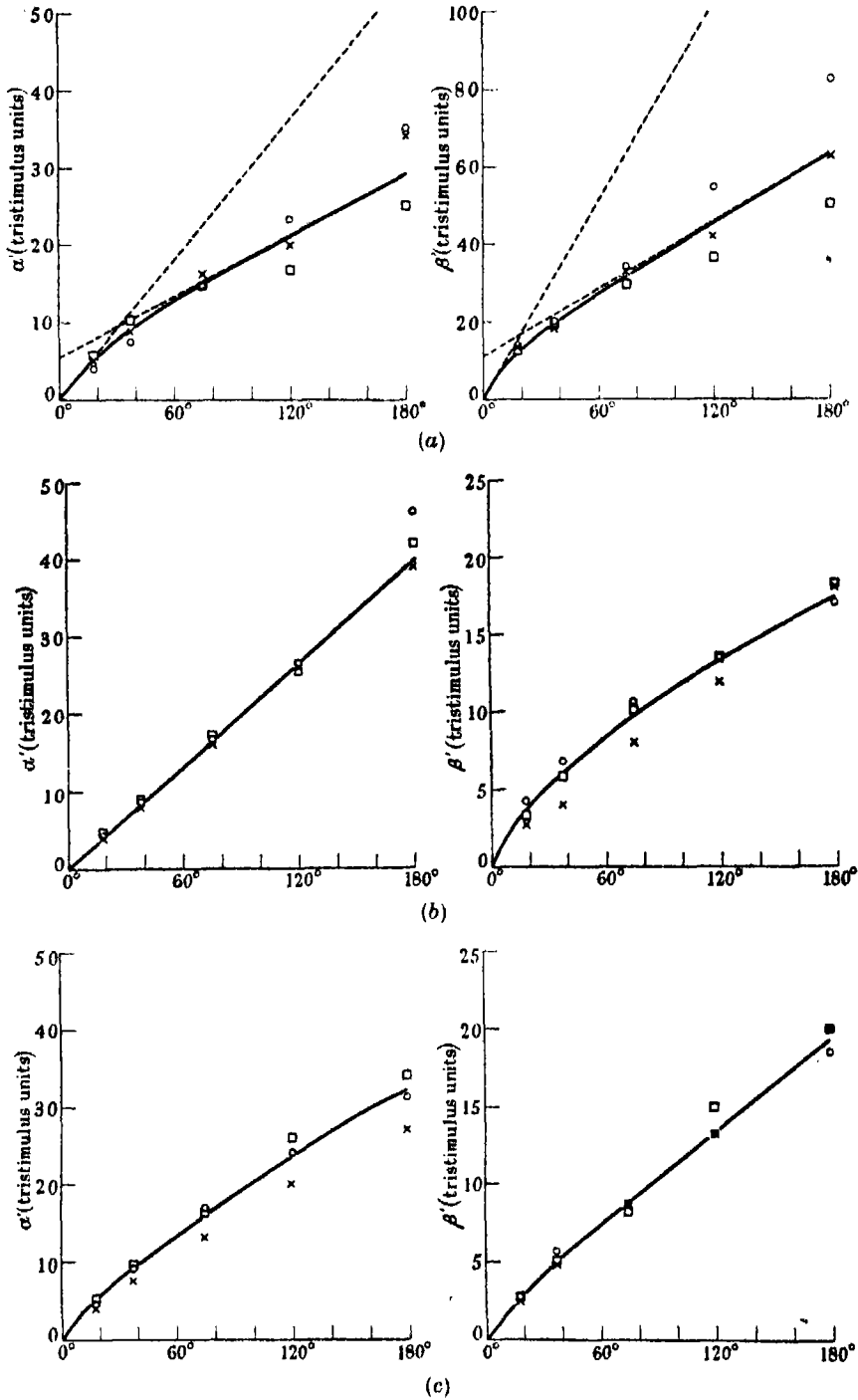


FIGURE 6. Typical curves showing relation between α' , β' , and test stimulus intensity. (a) Red adaptation 2000 photons, test stimulus 0.61 μ . (b) Red adaptation 2000 photons, test stimulus 0.65 μ . (c) Green adaptation 1200 photons, test stimulus 0.53 μ .

Consider the chain of events involved in the perception of light and colour. It is seen that there are three places where interaction may occur. Light enters the eye, is slightly absorbed in the optic media, passes through the inverted retina with its network of nerve fibres carrying the responses, into the receptors where the responses originate. These now pass across the anterior surface of the retina in the fibres and pass backwards down the closely packed optic nerve to the cortex. After this point we are not immediately concerned with the succession of events.

Interaction may take place, then, in the receptors, in the retinal nerve layers, or in the optic nerve.

At the risk of being too definite, we can imagine that the interaction in the receptors might be chemical, while that in the last two regions might be electrical in origin. The second is distinguished from the third, since it may be affected by the test stimulus itself in passing through to the receptors. Should this be so, one would expect the effect to be a function of wave-length and intensity, and the maintenance of a colour match of heterogeneous stimuli under different adaptation conditions argues fairly conclusively against this. For this purpose, then, the retinal nerve layers and the optic nerve may be regarded as a single unit, distinguished from the cones in which the supposed photoelectric and/or photochemical changes take place. The evidence obtained would not appear to be sufficient to decide which, or both, of the possible modes of interaction is responsible for the effect under consideration. One can only introduce ad hoc hypotheses, mentioning ' α -adaptation' (Schouten & Ornstein 1939), chain chemical reactions, and electrical induction, but in view of the prevailing uncertainty of our knowledge of retinal processes, it would be unwise in a communication of this nature to theorize further.

MODIFICATION OF RESULTS

In order to make new determinations of the fundamental sensations, it is necessary to apply corrections for the departure from the coefficient law, and we must decide how these corrections are to be made.

Referring to figure 6a it is clearly possible to have any constant of proportionality between test stimulus and matching stimulus, measured by the slope of the curve. This is the 'coefficient' of the generalized law, and there is no a priori reason for fixing it. There are, however, two values which seem more probable than any others. When the test stimulus is of a much higher intensity than the adaptation, the coefficient law holds good, so that if the curve of figure 6a could be produced indefinitely, it would eventually become linear. Adaptation to test and comparison patches is invoked for this condition, so that the 'coefficient' one would obtain by extrapolation at the high intensity end is likely to be seriously in error. If interaction only becomes negligible at high intensities, it is a seemingly insoluble problem to find the true 'coefficient'. A much more probable value is derived from the slope of the curve at very low intensities. It may be

that the interaction is a second order effect, negligible when the intensity is low, and relatively large at high intensities.

The derivation of corrected values of α and β is indicated graphically in figure 6a. Tangents are drawn to the curve and the vertical intercepts measured for a given horizontal traverse. The resulting measures can be inserted directly in equations (xiv), (xiva) and (xivb). It is found that the slope at high intensities yields values for H and J even more scattered and divergent than those previously obtained, while the alternative assumption is far more promising.

Consider a determination of H and J with red adaptation. It will be seen from figures 6a and 6b that with red adaptation the relative correction in the β 's is considerably greater than in the α 's. The first term in the numerator of (xiva) always turns out to be negative (i.e. b is positive), and increasing the β 's all by the same percentage makes this term numerically greater by a larger percentage, thus reducing the effects of random error. Actually with our convention, β'_1 is increased by more than β'_2 , so that this effect is still further increased. Moreover, the percentage increase in the first (squared) term under the square root sign is considerably more than in the second, and in all cases the change is found sufficient to make condition (xvii) satisfied, when with uncorrected values it is not. Another factor helping in the right direction, is the sensibly improved determination of the term $(\alpha'_1\alpha'_2 - \alpha_1\alpha'_2)$, since α'_2 has a bigger relative correction than α'_1 , and the condition for greater accuracy is satisfied.

Green adaptation had not yielded complex results, but the values of the fundamental responses derived from them are considerably altered by correcting the α' and β' values, the tendency being to agree better with the new determinations from red adaptation. In connexion with the apparently good agreement between the averaged results from red and green adaptation, the large spread of the individual determinations should be noted.

All the methods mentioned in the introduction give a value for the green fundamental J considerably removed from the spectral locus, while the red sensation H is often taken as that aroused by the longest visible wave-lengths, and my results are in general agreement. Additional circumstantial evidence for the approximate truth of the results, and the deductions I have made, lies in the facts mentioned in the account of the experimental results. I have supposed the responses to interact, and it is to be expected that when one is greatly preponderant it will be scarcely affected, while the other undergoes large perturbations. This is apparent in the red and green component recovery curves for 0.65μ under red adaptation, while since J is considerably beyond G from R , the green component recovery curve in 0.53μ exhibits the inhibition effect quite noticeably.

DISCUSSION OF ACCURACY

It is difficult to give a concise summary of this aspect of the work, since the range of errors is very large. One or two generalizations are possible, and we can say that on a normal day the maximum luminosity error for any one test stimulus is about

$\pm 20\%$, and when repetition is made and smooth curves drawn, the probable error is of the order of $\pm 5\%$, and on 'good' days somewhat better. The determination of the point at $t = 0$ on the recovery curve is of slightly greater accuracy than that of the later points.

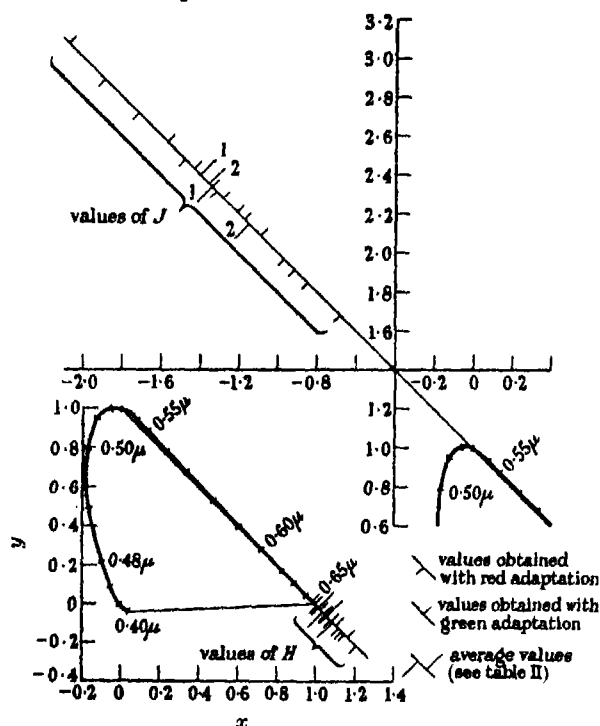


FIGURE 7. Corrected values of the red and green sensations H and J plotted in the W.D.W. trichromatic system.

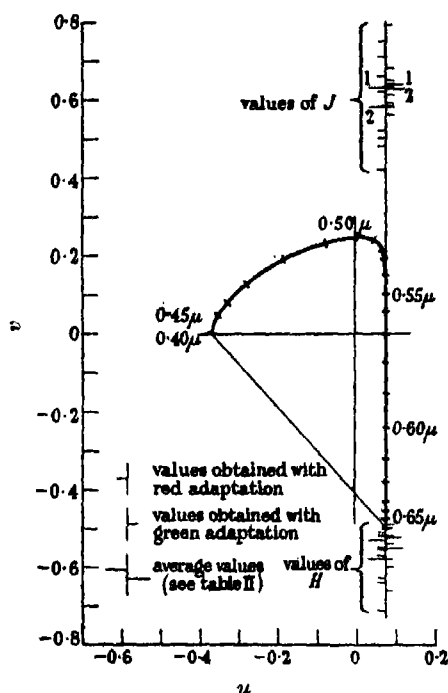


FIGURE 8. Corrected values of the red and green sensations H and J plotted in the R.U.C.S. diagram.

The main source of error lies in the use of test stimuli giving a ratio of α' to β' differing markedly from unity. The statistical spread in the smaller values may then be of the same order of magnitude as the quantities themselves. Here there occurs an impasse analogous to that confronting the atomic physicist in the Uncertainty Principle (Born 1935). To avoid stimuli which give binocular matches containing a relatively small amount of one of the primaries, the two test stimuli must be taken closer together, and an inspection of the working equations (xiii), (xiii a) and (xiii b) shows that the closer the stimuli the more accurate must the values of α' and β' be to preserve the same accuracy in the result for x/y . It is quite obvious from the limiting case when the two stimuli are made identical that there is nothing to be gained by having closely neighbouring test stimuli.

In spite of the improvements introduced by the corrections mentioned in the foregoing section, the terms $(\alpha'_1\alpha'_2 - \alpha_1\alpha'_2)$ and $(\beta'_1\beta'_2 - \beta_1\beta'_2)$ are easily the largest sources of error, the differences being small compared with the individual terms, and one of the measured values being subject to great uncertainty, i.e. α'_1 with red adaptation and β'_1 with green, the greatest accuracy being possible using the results at $t = 0$.

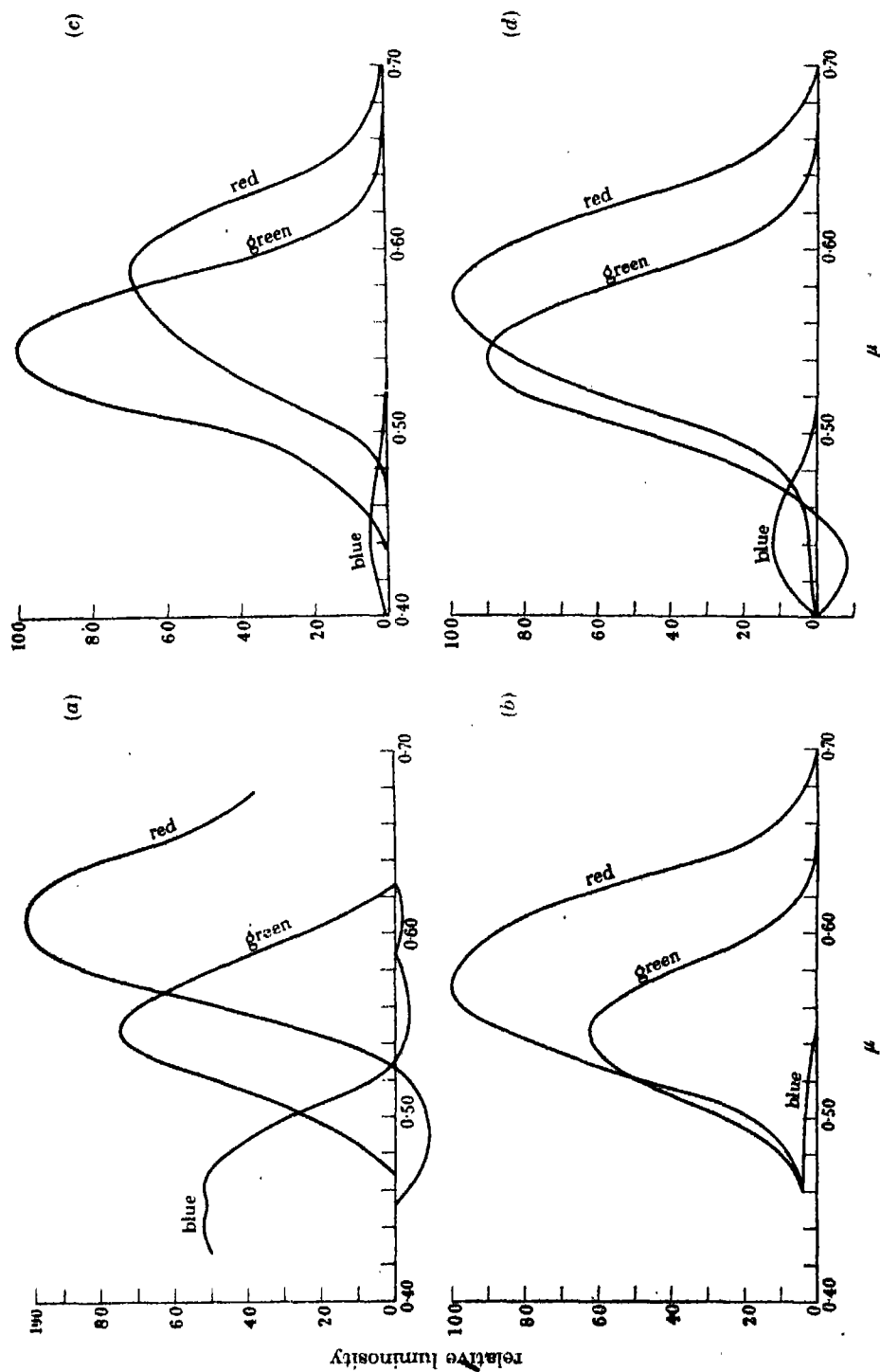


FIGURE 9. (a) 'Sensation' curves derived by Clerk Maxwell. These are of interest only in that they represent the first attempt at drawing sensation curves. (b) Abney's sensation curves. (c) König and Dieterici's curves corrected by D. B. Judd (1925). (d) Wright's sensation curves.

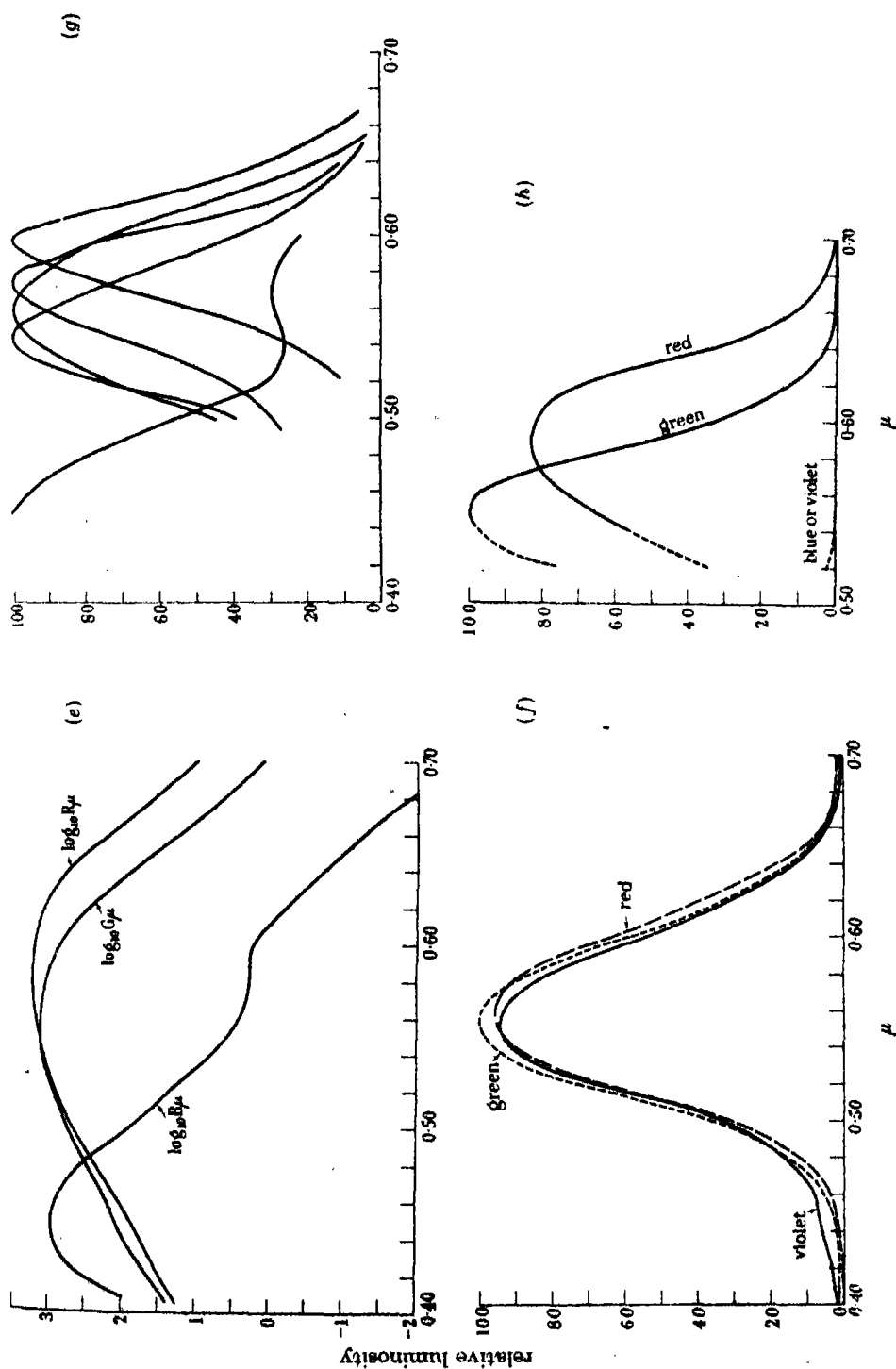


FIGURE 9 (*continued*). (e) Stiles and Crawford's curves plotted on a logarithmic scale. These were derived by a method involving threshold measurements, and their interpretation is not intended to be too strictly quantitative. (f) Hecht's sensation curves derived analytically to explain visibility and colour mixture quantitatively, and hue discrimination, saturation discrimination approximately. (g) Spectral sensitivities of elements from light-adapted frog's eye (Granit 1941. Reproduced by permission of the editor, *J. Opt. Soc. Amer.*). (h) Author's sensation curves for the wave-length range 0.54–0.70 μ . The approximate shape from 0.52 to 0.54 μ is indicated by a dotted line. Calculations based on the author's photopic luminosity curve (unpublished observations).

TABLE 2

recovery time sec.	adaptation colour and intensity	pair of test stimuli with intensities in photons (upper limit of 10:1 range given)	x_1/y_1	x_2/y_2	H				J			
					(W.D.W. system)	(W.D.W. system)	(W.D.W. system)	(R.U.C.S. system)	(W.D.W. system)	(W.D.W. system)	(R.U.C.S. system)	(R.U.C.S. system)
0	red, 2000 photons	0.61 μ (600) 0.625 μ (650)	- 17.6 - 102	- 0.46 - 0.57	1.06R - 0.06G 1.01R - 0.01G	1.06R - 0.06G 1.01R - 0.01G	- 0.85R + 1.85G - 1.32R + 2.32G	0.08U - 0.55V + 1.47W 0.08U - 0.51V + 1.43W	0.07U + 0.48V + 0.45W 0.07U + 0.62V + 0.31W	0.07U + 0.48V + 0.45W 0.07U + 0.62V + 0.31W	0.07U + 0.48V + 0.45W 0.07U + 0.62V + 0.31W	0.07U + 0.48V + 0.45W 0.07U + 0.62V + 0.31W
5		0.61 μ (600) 0.625 μ (650)	- 7.7 - 13.5	- 0.40 - 0.63	1.15R - 0.15G 1.08R - 0.08G	1.15R - 0.15G 1.08R - 0.08G	- 0.68R + 1.68G - 1.72R + 2.72G	0.08U - 0.64V + 1.56W 0.08U - 0.57V + 1.49W	0.07U + 0.42V + 0.51W 0.07U + 0.71V + 0.22W	0.07U + 0.42V + 0.51W 0.07U + 0.71V + 0.22W	0.07U + 0.42V + 0.51W 0.07U + 0.71V + 0.22W	0.07U + 0.42V + 0.51W 0.07U + 0.71V + 0.22W
10		0.61 μ (600) 0.625 μ (650)	- 5.8 - 17.7	- 0.49 - 0.60	1.21R - 0.21G 1.06R - 0.06G	1.21R - 0.21G 1.06R - 0.06G	- 0.97R + 1.97G - 1.48R + 2.48G	0.08U - 0.71V + 1.63W 0.08U - 0.55V + 1.47W	0.07U + 0.52V + 0.41W 0.07U + 0.66V + 0.27W	0.07U + 0.52V + 0.41W 0.07U + 0.66V + 0.27W	0.07U + 0.52V + 0.41W 0.07U + 0.66V + 0.27W	0.07U + 0.52V + 0.41W 0.07U + 0.66V + 0.27W
20		0.61 μ (600) 0.625 μ (650)	- 8.1 - 61.1	- 0.65 - 0.48	1.14R - 0.14G 1.02R - 0.02G	1.14R - 0.14G 1.02R - 0.02G	- 1.89R + 2.89G - 0.91R + 1.91G	0.08U - 0.63V + 1.55W 0.08U - 0.52V + 1.44W	0.07U + 0.75V + 0.18W 0.07U + 0.50V + 0.43W	0.07U + 0.75V + 0.18W 0.07U + 0.50V + 0.43W	0.07U + 0.75V + 0.18W 0.07U + 0.50V + 0.43W	0.07U + 0.75V + 0.18W 0.07U + 0.50V + 0.43W
0	green, 1200 photons	0.53 μ (380) 0.53 μ (380)	- 6.9 + 97	- 0.55 - 0.59	1.17R - 0.17G 0.99R + 0.01G	1.17R - 0.17G 0.99R + 0.01G	- 1.21R + 2.21G - 1.43R + 2.43G	0.08U - 0.60V + 1.58W 0.08U - 0.49V + 1.41W	0.07U + 0.59V + 0.34W 0.07U + 0.65V + 0.28W	0.07U + 0.59V + 0.34W 0.07U + 0.65V + 0.28W	0.07U + 0.59V + 0.34W 0.07U + 0.65V + 0.28W	0.07U + 0.59V + 0.34W 0.07U + 0.65V + 0.28W
5		0.53 μ (380) 0.53 μ (380)	- 21 - 34	- 0.61 - 0.52	1.05R - 0.05G 1.03R - 0.03G	1.05R - 0.05G 1.03R - 0.03G	- 1.57R + 2.57G - 1.09R + 2.09G	0.08U - 0.54V + 1.46W 0.08U - 0.52V + 1.44W	0.07U + 0.68V + 0.25W 0.07U + 0.56V + 0.37W	0.07U + 0.68V + 0.25W 0.07U + 0.56V + 0.37W	0.07U + 0.68V + 0.25W 0.07U + 0.56V + 0.37W	0.07U + 0.68V + 0.25W 0.07U + 0.56V + 0.37W
10		0.53 μ (380) 0.53 μ (380)	- 10 - 370	- 0.56 - 0.57	1.11R - 0.11G 1.00R —	1.11R - 0.11G 1.00R —	- 1.28R + 2.28G - 1.34R + 2.34G	0.08U - 0.60V + 1.52W 0.08U - 0.50V + 1.42W	0.07U + 0.61V + 0.32W 0.07U + 0.63V + 0.30W	0.07U + 0.61V + 0.32W 0.07U + 0.63V + 0.30W	0.07U + 0.61V + 0.32W 0.07U + 0.63V + 0.30W	0.07U + 0.61V + 0.32W 0.07U + 0.63V + 0.30W
20		0.53 μ (380) 0.53 μ (380)	- 22 - 12.5	- 0.67 - 0.54	1.05R - 0.05G 1.09R - 0.09G	1.05R - 0.05G 1.09R - 0.09G	- 2.08R + 3.08G - 1.17R + 2.17G	0.08U - 0.54V + 1.46W 0.08U - 0.58V + 1.50W	0.07U + 0.79V + 0.14W 0.07U + 0.58V + 0.35W	0.07U + 0.79V + 0.14W 0.07U + 0.58V + 0.35W	0.07U + 0.79V + 0.14W 0.07U + 0.58V + 0.35W	0.07U + 0.79V + 0.14W 0.07U + 0.58V + 0.35W
	red, 2000 photons	Average x_1/y_1 x_2/y_2 Average H, J	- 28 —	- 0.535 —	1.04R - 0.04G 1.08R - 0.08G	1.04R - 0.04G 1.08R - 0.08G	- 1.16R + 2.16G - 1.34R + 2.34G	0.08U - 0.53V + 1.45W 0.08U - 0.58V + 1.50W	0.07U + 0.58V + 0.35W 0.07U + 0.53V + 0.30W	0.07U + 0.58V + 0.35W 0.07U + 0.53V + 0.30W	0.07U + 0.58V + 0.35W 0.07U + 0.53V + 0.30W	0.07U + 0.58V + 0.35W 0.07U + 0.53V + 0.30W
	green 1200 photons	Average x_1/y_1 x_2/y_2 Average H, J	- 47 —	- 0.575 —	1.02R - 0.02G 1.06R - 0.06G	1.02R - 0.02G 1.06R - 0.06G	- 1.35R + 2.35G - 1.40R + 2.40G	0.08U - 0.52V + 1.44W 0.08U - 0.55V + 1.47W	0.07U + 0.63V + 0.30W 0.07U + 0.64V + 0.30W	0.07U + 0.63V + 0.30W 0.07U + 0.64V + 0.30W	0.07U + 0.63V + 0.30W 0.07U + 0.64V + 0.30W	0.07U + 0.63V + 0.30W 0.07U + 0.64V + 0.30W

Regarding the positive blue in the way indicated above (p. 35), and still assuming the red and green fundamental responses only to be aroused by the spectral region from about 0.70 to 0.54μ , we can safely neglect the blue at $t = 0$. The values of H and J are tabulated for various recovery times at which they were determined (table 2), and it will be seen that those at $t = 0$ are at least as consistent as those for later recovery times, and probably more so, though there are not sufficient observations to make a useful comparison of the spreads.

In figure 7 the values of H and J are plotted in the W.D.W. colour diagram, the most noticeable feature of the plot being that the J values are far more scattered than those of H .

A subject of current interest has been the derivation of a trichromatic system such that distances between points on the colour diagram bear a roughly linear relation to their colour difference, as evaluated by threshold measurements. So far the Rectangular Uniform Chromaticity Scale system (R.U.C.S. system) of Breckenridge & Schaub appears to possess most advantages (Breckenridge & Schaub 1939; Holmes 1940; Wright 1941). As a matter of some interest the values of H and J were transformed directly into the R.U.C.S. system using equations derived from data for the C.I.E. observer given by Smith & Guild (1932) and the transformation equations from the C.I.E. to R.U.C.S. system given by Holmes (1940). The selection of the R.U.C.S. system was an arbitrary process and its uniform chromaticity scales are only very approximate (Wright 1941), so there is no theoretical ground for thinking that the spreads of the two groups of points outside the spectral locus may be compared directly as an indication of the accuracy of the determinations. As was anticipated, however, the great disparity in the spreads of the values of H and J effectively disappear when they are plotted in the R.U.C.S. diagram (figure 8). The coefficients of H and J in the R.U.C.S. system are given in the last two columns of table 1 (uncorrected values) and table 2 (corrected values).

In the absence of a reliable value for K , the fundamental blue or violet sensation, one can only derive the red and green response curves for the region from 0.54 to 0.70μ , and their approximate shape from 0.54 to 0.52μ . These curves are given in figure 9. Doubtless the best test of the accuracy of the determinations will be in seeing the effect on the response curves of a small change in the values of H , J , or K .

I wish to thank Mr R. A. Lister and Mr T. H. Wang for their great assistance in recording my observations, and more especially Dr W. D. Wright, who not only did some recording and observing, but took a great interest in the work and exercised a watchful guidance that was never irksome. It is natural that with his knowledge of optics in general, and physiological optics in particular, he should have made a number of suggestions and given valuable hints. I must also thank the editor of the *Journal of the Optical Society of America* for permission to reproduce figure 9 (*g*). Finally, thanks are accorded to the Medical Research Council and the Swindon Education Committee for their financial assistance.

REFERENCES

- Abney, W. de W. 1913 *Researches in Colour Vision*. Sampson Low, Marston & Co.
 Born, M. 1935 *Atomic Physics*. Blackie.
 Breckenridge, F. C. & Schaub, W. R. 1939 *J. Opt. Soc. Amer.* **29**, 370.
 Clerk Maxwell, J. 1890 *Scientific Papers*, **1**, 410, ed. W. D. Niven. Camb. Univ. Press.
 Granit, R. 1941. *J. Opt. Soc. Amer.* **31**, 570.
 Hecht, S. 1932 *Report of a Joint Discussion on Vision*, p. 126. Phys. Soc. London.
 von Helmholtz, H. 1911 *Physiologischen Optik*, **2**, 119 (3rd ed.).
 Holmes, J. G. 1940 *Proc. Phys. Soc.* **52**, 359.
 Judd, D. B. 1925. *J. Opt. Soc. Amer.* **10**, 635.
 König, A. 1929 *Handb. Exp. Physik*, **20**, 105.
 von Kries, J. 1904 *Nagel's Handb. Physiol. Menschen*, 211.
 Parsons, J. H. 1915 *Colour Vision*, sect. 3, chap. 1.
 Peddie, W. 1922 *Colour Vision*, chap. 3.
 Schouten, J. F. & Ornstein, L. S. 1939 *J. Opt. Soc. Amer.* **29**, 188.
 Smith, T. & Guild, J. 1932 *Trans. Opt. Soc.* **33**, 73.
 Stiles, W. S. 1939 *Proc. Roy. Soc. B*, **127**, 64.
 Stiles, W. S. & Crawford, B. H. 1933 *Proc. Roy. Soc. B*, **112**, 428.
 Wright, W. D. 1929 *Trans. Opt. Soc.* **30**, 141.
 Wright, W. D. 1934 *Proc. Roy. Soc. B*, **115**, 49.
 Wright, W. D. 1939 *J. Sci. Instrum.* **16**, 10.
 Wright, W. D. 1941 *Proc. Phys. Soc.* **53**, 93.

The selection of invisible mutations

BY K. MATHER AND L. G. WIGAN

John Innes Horticultural Institution, Merton, London

(Communicated by C. D. Darlington, F.R.S.—Received 5 March 1942)

An inbred line of *Drosophila melanogaster* was selected, in two separate experiments, for the manifestation of polygenic characters over periods of twenty-one and fifty-three generations respectively.

In one case change with selection, when it occurred, was smooth, but in the second experiment it proceeded in the form of sudden steps separated by periods in which selection was ineffective.

These jumps are due to the recombination of the mutations to which individual polygenes have given rise. Masked by the non-heritable fluctuations of the character, these mutations, having small individual effects, accumulate until recombinations give rise to more extreme variants, which fluctuation can no longer hide. Selection then becomes effective in producing a change in the character.

1. SELECTION, VARIATION AND MUTATION

One of the conditions of stability in the genetical constitution of a population of organisms is that the relative frequencies of the various allelomorphs of any gene shall be constant. Conversely, with some minor exceptions, changes of genetical constitution all involve changes in the relative frequencies of the allelomorphs at one or more loci.

Such changes may be brought about in two ways. In the first place chance fluctuations of allelomorphic frequencies may occur; for there are always considerably fewer zygotes surviving to the breeding stage in any generation than there were gametes of either kind produced by the previous generation. Hence each generation of zygotes is in effect a sample from the much larger number that would have occurred if all the gametes of their predecessors had functioned. As a consequence the allelomorphic frequencies must be subject to sampling variation. Fluctuations of this type do not usually, however, have any lasting effect, for, in general, any allelomorph has an equal chance of spreading or becoming more restricted in any generation. Either process is thus potentially reversible. But the process is irreversible if an allelomorph is completely lost, unless, of course, that same allelomorph should recur *de novo*. The only stable change is a reduction in allelomorph frequency, and there is always a finite probability of such a random loss of allelomorphs. The probability of occurrence of such loss is inversely proportional to population size (Fisher 1930).

✓The second agency which may change the allelomorph frequency is selection. If, as a result of its action on the phenotype, one allelomorph enjoys an advantage over the others at the same locus, its frequency will, on the average, grow at the expense of those of its rivals, until finally it completely replaces the others. Once again the change has led to a reduction in the number of allelomorphs existing in the population. ✓

Selection need not always result in the loss of allelomorphs, for sometimes it favours a heterozygote at the expense of both corresponding homozygotes (Fisher 1939). In such cases selection is acting to stabilize allelomorph frequencies. It is then an agent of preservation rather than an agent of change; but wherever progressive changes go on under the action of a selective force changes in allelomorph frequencies will be found to occur and hence, in the long run, allelomorphs will be lost.

It is not difficult to see that changes in the genetical constitution of a population are impossible unless there exists heritable variation, i.e. unless at least one gene is represented by two or more allelomorphs. So the effective action of selection depends on, and at the same time destroys, heritable variation. In other words, selective changes must eventually bring about genetical homogeneity in the absence of any counter-balancing agent. Such an agent is provided by point mutation, the spontaneous origin of one allelomorph from a different pre-existing allelomorph. It is to be presumed that on the average the mutation process replaces variability at approximately the same rate as it is being destroyed by selection and random fluctuation. ✓

The existence of a mutation process has long been recognized, and mutations have been observed to occur in all organisms which have been subjected to intensive genetical investigation. It has, however, been almost inevitable that the recognition and study of mutations should be confined to cases of genes having marked individual effects, such as lethality, gross morphological change or altered coloration. Indeed, it is only by the study of such genes that the occurrence of point mutation

could be observed and verified. These major changes are, however, of more importance in experiment than in nature.

A different type of character also exists whose inheritance is controlled by many genes each having an effect which is small compared with non-heritable variation and roughly equal to those of the other genes. Characters of this kind, termed polygenic or quantitative, seem to have great importance in adaptation and evolution. Little attempt has, however, been made to study the mutation of these polygenes. Each has such a small effect when compared with environmentally determined fluctuations that direct study by the methods used for major genes is impossible. But these difficulties are of little importance, for, inasmuch as selection acts on the total amount of heritable variation available, our interest must be directed more towards the total amount of new variation of any character, than towards the behaviour of the individual polygene. Knowledge of the behaviour of single polygenes would be of little help unless their number were also known, and there seems to be little prospect of such information being obtained. It is of greater importance to observe the origin of all new variation in the character.

The preliminary step to detecting and measuring the rate of origin of new heritable variation is to build up a line which is homozygous, or very nearly so, for the polygenes affecting the character in question. Such a line would show little phenotypic variation, unless the environment were playing a big part, and in any case would not respond to selection. As new genetic variation arose the phenotype would be more variable and selection could effect changes in the genotype. East (1935) has described a case in *Nicotiana* where a homozygous line, when newly arisen by parthenogenesis and somatic doubling, was almost invariable phenotypically, but which began to show considerable variation after a few generations, during which mutation might be supposed to have occurred.

In general, however, response to selection affords a better test of the existence of variation, and most investigations have been conducted on this basis. Two deserve special mention as bearing on the question of polygenic variation. Johannsen (1909), in his classical experiments with inbred beans, selected for increase and decrease in the manifestation of various characters over a period of six years, but no detectable change took place. His beans were homozygous or nearly so when the selection began, and the rate of mutation was not sufficiently great to provide any effective amount of new variation during the period of the experiment. Lindstrom (1941) made a fully homozygous diploid tomato by somatic doubling from a haploid, and later selected its descendants for decreased vigour. After nine generations a slight fall in fruit size and seedling weight was detectable in some, but not all, of the lines. It would thus seem that new variation appears very slowly by mutation and that large numbers of generations must be raised before its effects become marked. This can be done with a rapidly breeding organism such as *Drosophila melanogaster* in a way which is impossible in annual plants.

The behaviour of polygenic characters in *D. melanogaster* has been described by Mather (1941) and Wigan (1942). In their experiments the material on which

selection was exercised originated as hybrids between two laboratory stocks and hence was of a highly heterozygous nature. One of the parental lines had, however, been inbred over a long period and so afforded an excellent opportunity for a study of the origin of polygenic variability by mutation. It is with the effect of selection on this line that the following account is concerned. Selection was originally started as part of the analysis of the interstock hybrids, and the results obtained in the first few generations have already been described by Mather and Wigan in the papers cited above, but they will be included here with the later generations in order that a comprehensive account of the experiments may be given.

2. THE EXPERIMENTAL TECHNIQUE

Detection of the variation originating by mutation depended on the response of the material to selection for increased and decreased manifestation of the characters concerned. The flies with which the experiments started were very highly inbred and so were expected to be nearly homozygous (Haldane 1936) and to show no reaction to selection. Thus any later response could be attributed to the effects of mutation in introducing variation upon which selection could act.

The flies came from the inbred Oregon + stock. When selection commenced, this line had been kept in the laboratory by brother-sister mating for nineteen generations. Previously it had been brother-sister mated for fifty-nine generations by Dr U. Philip of University College, London, to whom we are indebted for the stock. During these seventy-eight generations some mutations had presumably occurred, but the consequent variation would be eliminated by the mating system, for all the parents of any generation were themselves the offspring of a single pair. A small amount of such variability might, of course, be expected in any generation as a result of recent mutation, but that its effects were negligible is shown both by the failure to respond to selection in the early generations of the experiment and by a special regression analysis which was undertaken prior to the commencement of selection. It was observed in this analysis that the degree of manifestation of polygenic characters in a progeny was independent of the manifestation in the parents, i.e. heritable variation was very small or absent. It may be remarked that some variation must have been eliminated by the mating system during the course of the experiments, for, though inbreeding was relaxed, each generation included not more than 120 flies. Variant types would, however, be exposed to selection between the mutation and chance elimination. Thus advances with selection are expected to occur.

Two polygenic characters were used. One of these, the number of chaetae on the ventral surface of the fourth and fifth abdominal segments was used by Mather (1941), and the other, the number of sternopleural chaetae, by Wigan (1942), in their analyses of response to selection in hybrids. Separate experiments were made on these characters. In the first, when the sternopleural number was employed, three single-pair matings were made in each generation. During the first two

generations no selection was practised, the parents being taken at random. Afterwards, however, selection for both increased and decreased chaeta number began. In the high line the male and female with the greatest numbers of chaetae, and in the low line the male and female with the smallest numbers of chaetae, were chosen from each progeny to be parents of the next generation. Brother-sister mating was never practised where avoidable, the male from progeny 2 being mated to the female from progeny 1, the male from progeny 3 to the female from progeny 2, and the male from progeny 1 to the female from progeny 3, throughout the major part of the experiment. Two departures from this rule must be mentioned. In the first three selected generations the most extreme female was mated to the more extreme of the two males from the other progenies, and so on. Secondly, in some generations one or two cultures failed. If one failed two extreme females from one successful culture were mated to the two extreme males from the other culture, with one mating in the reciprocal direction. On the very rare occasions when only one culture succeeded, three brother-sister matings had to be made, but these cases were confined to the end of the experiment. The mating system used serves to conserve any new variation and to give the mutant genes, which may arise in different cultures, full opportunity of recombining.

This first experiment was conducted throughout its twenty-three generations in tubes. At first a single pair was allowed to lay for two successive periods of 24 hr. in two tubes. Later, when some reduction of fertility was observed, these periods were extended to 48 hr., but in consequence of the overcrowding to which this procedure gave rise, the laying periods were soon reduced to 40 hr. in the first tube and 30 hr. in the second. A few sample counts showed that flies from the two tubes, in which any given female had laid, had the same mean chaeta number, and so the two sets of flies were pooled for the purposes of record and selection. Twenty flies of each sex were counted from each progeny unless the total yield fell short of that number.

Selection was based on the number of abdominal chaetae in the second experiment. Here the flies were raised in half-pint milk bottles. Three single-pair matings were used in each generation of each line, there being, as before, two lines, one selected for increased and the other for decreased chaeta number. The extreme male and female were chosen from each progeny and, with the exception that brother-sister matings were avoided, the most extreme of the three selected males were mated to the most extreme female and so on. When one, or very rarely two, progenies failed the procedure was the same as in the first experiment.

The single pairs of flies were mated in tubes and later transferred to standard bottles, in which they remained for 5 days. This led to little if any overcrowding, especially in the later generations when fertility fell away. As before, twenty flies of each sex were counted from each progeny, where available.

In both experiments corresponding generations of high and low lines were raised side by side in the incubator in order that any environmental influences should affect the two as nearly equal as possible.

3. THE RESULTS OF SELECTION

The results of the first experiment are set out in table 1 and shown graphically in figure 1. The generations before selection began require little comment, though it would appear that the original Oregon+ stock which was raised in bottles had a slightly lower chaeta number than its unselected descendants raised in tubes. This could reasonably be attributed to the different cultural conditions prevailing in bottle and tube.

TABLE 1. MEAN NUMBER OF STERNOPLURAL CHAETAE OF LINES OF INBRED OREGON+ FLIES SELECTED FOR HIGH AND LOW CHAETA NUMBER

selected generation	females		males		compound
	high line	low line	high line	low line	
	unselected	20.05 18.90 19.70	19.73 18.90 18.93		
1	20.15	18.95	19.73	19.15	+ 0.89
2	19.57	19.80	18.95	19.20	- 0.24
3	19.90	19.40	19.57	18.90	+ 0.58
4	19.67	19.58	19.58	18.83	+ 0.42
5	19.98	19.55	20.30	19.18	+ 0.78
6	19.13	18.73	18.45	18.32	+ 0.27
7	19.66	18.93	19.07	19.00	+ 0.40
8	20.05	19.70	19.97	19.35	+ 0.49
9	20.28	18.62	19.73	19.25	+ 1.07
10	19.65	19.33	20.00	19.07	+ 0.62
11	19.94	18.95	19.27	18.60	+ 0.84
12	19.42	18.70	19.09	18.25	+ 0.78
13	19.85	18.67	19.80	18.13	+ 1.43
14	19.72	18.53	19.35	18.30	+ 1.12
15	19.58	18.67	18.80	17.91	+ 0.90
16	20.18	19.17	19.45	18.60	+ 0.93
17	19.38	18.10	19.65	18.55	+ 1.19
18	19.69	18.95	19.94	18.93	+ 0.87
19	19.72	18.28	19.62	18.38	+ 1.34
20	19.62	19.38	19.90	18.91	+ 0.62
21	20.03	18.41	20.00	18.73	+ 1.45

The first few generations of selection appeared to give little consistent effect. Indeed, in the second of these the low selection line had more chaetae than the high line. It soon became clear, however, that the two lines were diverging in mean chaeta number. This is well shown by the differences between high and low lines given in the last column of table 1. These figures are obtained as the mean of the two differences, one from each sex, between the mean chaeta numbers of the two lines. They are thus compound measurements which represent the general difference between the two selection lines in any generation, and they should be less subject to environmental fluctuations than the line means, because the differences are taken between chaeta numbers of flies raised simultaneously in the incubator.

These compound differences show some fluctuation, especially in the earlier generations, but the upward trend is steady and unmistakable. After twenty-one generations of selection the high and low lines differed by an average of between 1.25 and 1.50 chaetae per fly, and this difference has developed gradually over the whole course of the experiment. Figure 1 shows, however, that the high line had been stable, the change having been confined to the low line, where mutations have occurred and were selected steadily over the whole time of the experiment.

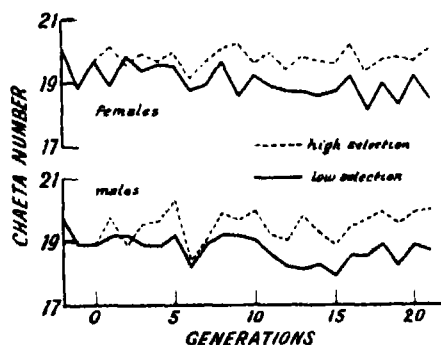


FIGURE 1. Twenty-one generations of selection for sternopleural chaeta number in *Drosophila melanogaster*. No selection was practised until the third generation.

The results of the second experiment in which abdominal chaeta number was selected are given in table 2 and figure 2. The behaviour of the flies under selection is more complicated in this than in the first experiment. The low selection line was kept in being continuously throughout the fifty-three generations of the experiment, and from figure 2 it appears that no change occurred until generation 24. Prior to that generation marked fluctuations had been observed but no steady change with selection was detectable. Between generations 24 and 29 there was a marked decrease in the mean chaeta number of this line. It is difficult to judge the precise magnitude of this fall, because of the disturbing effect of the fluctuations, but it is certainly about 1.5 chaetae. After this relatively rapid fall a period of stability set in and lasted until somewhere about generation 42. A second response to selection of about the same magnitude then began and was probably concluded by generation 47, though again fluctuation renders it difficult to be quite sure of the precise end-point. The main point is, however, clear, viz. that the low line reacts to selection in the same way as hybrid material, by a series of alternating periods of rapid change and lengthy stability. The significance of this will be discussed later.

The high selection line was carried on in a slightly different way. It showed no change with selection until generation 13, when a period of increase set in. By generation 17 the high-line mean exceeded that of the low-line by about 1 chaeta. It may be noted that in the males there was apparently a slow but steady increase in the early generations, but as the same trend appeared in the low-line males too

it must be attributed to non-heritable causes. Thus the high line shows the same type of behaviour as the low line, initial stability followed by a distinct period of change.

At generation 17 the original high line was discontinued and a new one started by back selection from the low line, which, it will be recalled, had shown no phenotypic change up to that time. At first this new high line reproduced the behaviour of the low line from which it was taken, but after generation 23, i.e. after six generations of selection, it began to depart from its progenitor. The total departure

TABLE 2. MEAN NUMBER OF CHAETAE ON THE FOURTH AND FIFTH ABDOMINAL SEGMENTS OF LINES OF INBRED OREGON + FLIES SELECTED FOR HIGH AND LOW CHAETA NUMBER

selected generation		females		males		compound
		high line	low line	high line	low line	
	parental	42.00	42.00	38.24	38.24	
1		44.46	43.08	38.39	38.17	+ 0.80
2		42.64	43.68	38.17	38.24	- 0.56
3		43.21	43.25	37.73	38.45	- 0.38
4		44.35	42.67	38.72	38.00	+ 1.20
5		43.36	44.35	38.03	36.53	+ 0.27
6		43.93	43.53	38.37	38.50	+ 0.14
7		43.05	43.63	38.96	38.87	- 0.24
8		42.65	42.40	38.40	38.65	0.00
9		44.67	44.48	38.90	39.13	- 0.04
10		44.17	43.55	38.77	37.83	+ 0.78
11		43.52	43.40	39.12	39.45	- 0.11
12		43.48	44.17	39.45	39.15	- 0.20
13		42.88	42.67	38.20	38.48	- 0.03
14		44.96	44.03	38.97	38.55	+ 0.69
15		45.47	43.68	39.34	38.87	+ 1.13
16		45.47	43.23	39.48	38.85	+ 1.43
17		43.95	42.67	39.40	38.47	+ 1.13
second high low						
1	18	43.68	43.18	38.02	38.35	+ 0.09
2	19	44.15	43.63	39.12	37.95	+ 0.85
3	20	42.72	43.60	38.80	38.95	- 0.51
4	21	43.77	42.23	38.10	38.30	+ 0.62
5	22	44.48	43.05	38.63	37.90	+ 1.09
6	23	43.77	43.65	38.58	39.10	- 0.32
7	24	44.90	43.85	39.68	38.83	+ 0.96
8	25	44.42	42.80	38.87	38.30	+ 1.09
9	26	43.95	42.65	38.55	38.43	+ 0.72
10	27	44.78	43.00	39.97	37.63	+ 2.06
11	28	44.62	42.12	39.67	37.20	+ 2.48
12	29	44.67	40.40	39.55	35.95	+ 3.94
13	30	45.93	42.15	40.57	36.85	+ 3.76
14	31	45.75	41.40	40.18	36.62	+ 3.96
15	32	46.23	43.37	41.83	37.22	+ 3.74
16	33	47.91	42.65	42.91	36.55	+ 5.86
17	34	47.20	40.35	43.41	36.58	+ 6.64

TABLE 2 (continued)

selected generation		females		males		compound
		high line	low line	high line	low line	
		42.00		38.24		
parental						
third high	low					
1	35	42.64	42.35	36.49	36.73	+ 0.03
2	36	41.95	42.60	37.25	36.86	- 0.13
3	37	43.30	41.40	38.15	35.95	+ 2.05
4	38	43.50	42.45	37.48	36.10	+ 1.22
5	39	43.48	41.85	37.30	36.90	+ 1.02
6	40	43.03	42.25	37.25	37.63	+ 0.70
7	41	42.58	41.53	37.28	35.53	+ 1.40
8	42	44.10	39.65	37.20	37.10	+ 2.23
9	43	43.48	40.95	37.80	35.60	+ 2.37
10	44	44.53	40.94	39.17	36.16	+ 3.30
11	45	43.87	43.00	39.38	35.95	+ 2.15
12	46	43.52	40.52	39.30	35.30	+ 3.50
13	47	43.82	40.15	38.05	34.65	+ 3.54
14	48	42.97	40.08	37.87	34.00	+ 3.38
15	49	44.48	39.70	38.50	34.88	+ 4.30
16	50	43.78	38.75	38.03	34.17	+ 4.45
17	51	44.92	41.00	39.37	35.46	+ 3.92
18	52	44.25	40.70	38.04	34.65	+ 3.47
19	53	44.05	40.34	38.88	34.25	+ 4.17

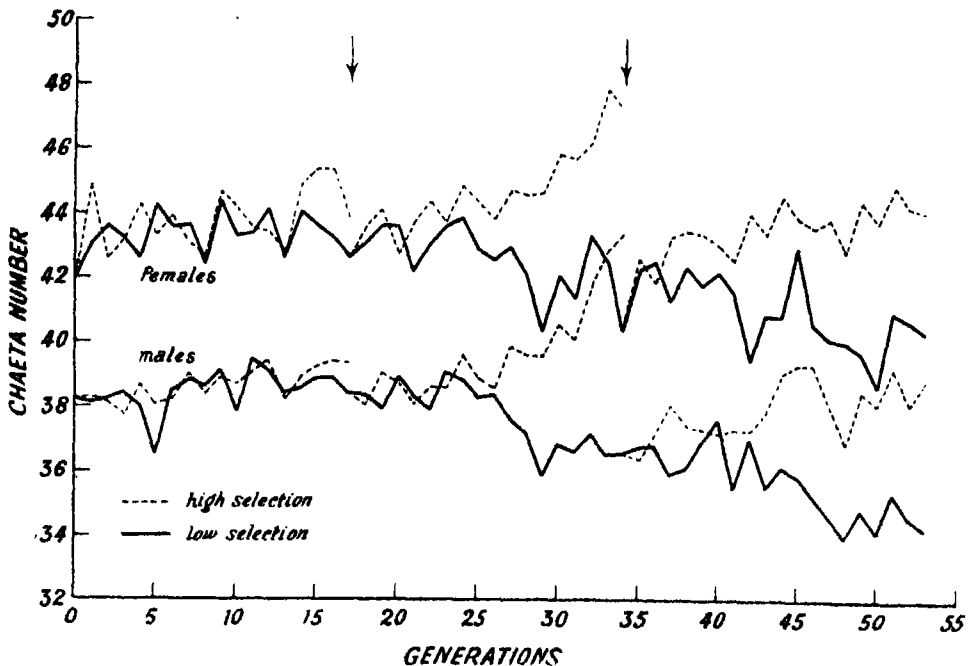


FIGURE 2. Fifty-three generations of selection for abdominal chaetae number in *Drosophila melanogaster*. Three high-selection lines were used, the second and third being back selections taken from the low line at the places marked by arrows.

developed very rapidly as the low line began to fall at about the same time, but figure 2 leaves no doubt that the high line was itself increasing rapidly during this period. This second line was itself terminated at generation 34 after about ten generations' increase, during which time its mean had changed by roughly four chaetae. The compound difference in the last column of table 2 suggests a short period of stability between generations 29 and 32, followed by a second rise, but this must be treated with some caution as both high and low lines were involved in changes about that time. Such a period of stability is, however, not unlikely, and if real serves to emphasize the jerky nature of change under the action of selection.

A third high line was started by a new back selection from the low line at generation 34, i.e. during the second period of stability in the low line. This new line almost at once showed a response to selection and in four generations had recovered all the ground lost by the parental low line through selection in the opposite direction. Though subsequent progress was slow, it appeared to be changing to some extent right up to the end of the experiment, i.e. for another twelve generations or so. By this time its mean had advanced slightly beyond that of the unselected material with which the experiment was started.

The high line was thus reinitiated twice by back selection from the low line. On each occasion progressive increase in chaeta number began earlier than in the preceding high lines. Hence it appears that the low line itself was changing by accumulating variations which made possible more rapid increases in chaeta number.

4. MUTATION AND THE RESPONSE TO SELECTION

The two characters chosen as the bases for selection in the experiments usually show polygenic variability. The characters may show any degree of expression between fairly wide limits, are subject to environmental effects which are not small when compared with genetic differences, and, though showing heritable variation, rarely permit division into classes marking simple segregation. In hybrids they behave in a polygenic manner, and in any case genes controlling the character have been shown to occur in each major chromosome (Mather 1942). This, however, does not mean that single major mutations never affect the characters. Genes like *scute* and *Hairless* drastically reduce the number of chaetae in both positions; but these major effects are easily distinguishable from the polygenic variation which is normally observed in both characters.

The question then arises as to how the results of the selection experiments, especially the second one, should be interpreted. Are we to suppose that the new variation arose by mutation of a number of polygenes with relatively small effects, or are we to attribute the changes to selection causing single major mutations to spread throughout the experimental material?

In the second experiment, at least, the latter alternative appears at first sight to have much to commend it. Progress was made in a series of jumps, each of

which occupied about five generations, i.e. about the time a single major change would require to spread through the material under the system of mating used in the experiments. A closer examination, however, brings to light serious difficulties in this interpretation. The low line of the second experiment showed two periods of change and the various high lines at least three, and perhaps four, between them. In other words five or six genes are involved, unless we are willing to suppose that the same gene mutated several times. Even then two genes are involved as it would appear unreasonable to attribute to the same gene changes towards reduced number in the low line and towards increased number in the high line. If five or six genes are admitted, the basic distinction between this view and the alternative polygenic interpretation is rapidly disappearing, and can only be maintained artificially.

If, on the other hand, the changes are attributed to the action of two fairly labile genes certain new difficulties arise, for the properties of the various changes with selection are not all alike. The first high line was derived from the same unselected parents as the low line, and the second high line by back selection from the low line after seventeen generations. During these seventeen generations the low line had failed to respond to selection, that is, on the view we are discussing, had failed to show any mutation. Yet the second high line began to depart from the parental value sooner and made greater progress than did the first high line. We can attribute this to chance, perhaps, but when it is seen that the third high line reacted to selection still sooner and still more rapidly, this explanation becomes somewhat strained. Superficially it can be saved by supposing that the onset and rate of spread of the new variants differed on the various occasions by reason of the operation of unlike conditions. But the environment had not materially altered, nor had the mode of selection been revised, and so these altered conditions must be supposedly genetic. The discussion is then back to the starting point and the polygenic view has not been avoided.

There is one further point which argues against the view under discussion. The character in question has been shown to behave polygenically in hybrid material (Mather 1941, 1942). If polygenes are known to affect the number of abdominal chaetae, where lies the advantage in trying to account for new variation on the basis of single major changes? Such an attempt would imply that the polygenes had a remarkably low mutation rate, in that it attributed none of the new variation to a change in any one of these presumably numerous genes. This cannot be proved wrong, but it is a very large assumption to make for the support of an hypothesis which is weak on other grounds.

Finally, in the hybrid material alternating periods of stability and rapid change were observed and shown to be due, not to mutation, but to recombination of polygenes. The behaviour of the new variation is open to the same interpretation, especially as the changes observed are neither so frequent nor so large as in the hybrids, which would obviously have more polygenic variation available.

To sum up, the view that attributes the response to selection, in the second experiment, to single relatively large mutations meets with difficulty when a

detailed analysis is made. It can be saved only by a series of progressively less likely ad hoc supporting hypotheses.

The only difficulty in accounting for the results on the hypothesis of polygenic mutation lies in the fact that response to selection, at least in the second experiment, was not smooth but took place in several rapid steps separated by long periods of stability. The new variation must have been accumulating steadily as mutations occurred, and yet it became available to the action of selection in short bursts.

We have seen above that similar behaviour in hybrid material was attributable to recombination. Polygenic combinations can be balanced in the sense that they include both + and - genes in roughly equal numbers. They become unbalanced when recombination causes this approximate equality to break down. In hybrid material effective recombinations occur freely, and the ensuing unbalance is such as to give large changes in the expression of the character controlled by the polygenes. Where the variation is new and arises by mutation the effects of recombination should be seen less frequently and be considerably smaller in magnitude. This is in fact the case as a comparison with the earlier results (Mather 1941) will show. The effective recombination then occurred within four or five generations and gave selective advances of four or even eight chaetae. In the present experiment the first response to selection developed after thirteen generations and consisted of a mean advance of just over one chaeta.

Evidence of the accumulation of variability prior to the occurrence of a selective advance, i.e. while the line is still apparently stable, is afforded by the comparative behaviour of the various high selections. The first of these was initiated simultaneously with the low line from the highly inbred unselected Oregon + stock. It showed no advance for thirteen generations and then increased by rather more than one chaeta in three or four generations. After this it appeared to have become stable again. The low line during this period showed no phenotypic change. The second high line originated in a back selection from the seventeenth generation of this unchanged low line, but began to show a movement with selection after only six or seven generations and its advance was much greater than that of the preceding high line. This indicates that the apparently unchanged low line had in fact acquired a store of hereditary variability such as would give advances in the high direction. It may be remarked that the low line would then be expected also to have acquired a store of variability of the kind which would give advances in the low direction. The effect of this variability was seen in the drop beginning in generation 24 or 25.

The third high line differed slightly from its predecessors. It was back selected from the low line after the latter had itself responded to selection. The response in the high direction set in immediately and soon restored the line to the level of the original unselected stock. Progress was afterwards slower but fairly definite. This shows that the response of the low line had not used up all the variability originating by mutation. Some remained after the low response had ceased and

was immediately available to high selection. Since recombination has the effect of unbalancing combinations in opposite directions simultaneously, such a result is to be expected.

Similar tests were not made in the first experiment, but the fact that in it high selection was inoperative, while the low selection had an effect, would indicate that a similar process was going on. Mutation must have happened in both lines and there is no reason to suppose that mutants were all directed towards causing lower sternopleural chaeta number. Hence it seems likely that the high line was in fact acquiring variability, though its phenotype had not changed.

Two questions now arise, viz. how the genic variability arising by mutation came to be organized so that it was freed by recombination, and what agent was responsible for protecting the new variation from selection during the time when it was accumulating. The answer to the former question is not difficult to find. When one mutation ($A \rightarrow a$) has occurred, the new allelomorph, say for the sake of argument, tending to decrease chaeta number, and both allelomorphs have survived in the population, a second similar mutation ($B \rightarrow b$), if in the same chromosome, has at best a half chance of arising in that representative of the chromosome which contains a . If a has not spread to any extent in the population, b is more likely to occur with A . Recombination is then necessary to bring a and b together and so give the maximum phenotypic variation. With a greater number of mutations the chance of them arising in such a way as to give the most extreme phenotype is still smaller, and recombination comes to play an even more important part. In other words, mutation, in not being directional, tends most frequently to give partially balanced combinations. These are then broken down by recombination and the bursts of response to selection are brought about.

The answer to the second question regarding the agent protecting variation from selection is perhaps less certain; but there are definite indications of its nature. Both the phenotypic characters on which selection was based are subject to marked fluctuations of a non-genetic nature. Now such fluctuations lower the efficiency of selection in picking out hereditary variation. Individuals may be chosen as parents of the next generation largely or wholly by virtue of their non-heritable departure from the culture mean, and so genotypic variation will persist under cover of these selected fluctuations. The greater the magnitude of the fluctuations relative to the genetic variation, the less effective will selection of a given rigour be in changing the mean expression of the line and in reducing the genetic variation. This damping effect of fluctuation must be at its maximum where mutations are apparently the only source of heritable variation. The latter will thus accumulate under the cloak of fluctuation. Once, however, recombination gives rise to a combination of polygenes of large effect this will be selected in spite of fluctuations, and, in spreading through the population, will give a selective advance. The conditions of the experiments were such as to maximize this damping effect, for selection was always based on phenotypes rather than on breeding behaviour, very small populations were used and inbreeding was relaxed.

Evidence supporting this view is obtainable from the relative behaviour of the flies in the two experiments. The first one showed a much smoother response to selection, though, as we have seen, there is reason to believe that mutational variation was accumulating in the high line. We should, however, expect to find that sternopleural chaeta number was less subject to fluctuation than the abdominal chaeta number of the second experiment.

A number of flies had both types of chaetae counted at various times in the first experiment, five of each sex being taken from each culture treated in this way. The pooled variances of these flies round their culture means are given for each character in table 3. The abdominal variances are more than three times as big as the corresponding sternopleural values. What allowance should be made for the difference in mean number is difficult to judge, but it does appear that even after a reasonable correction the sternopleurals were less variable. Since the variances are based on intraculture comparisons they measure fluctuations to the almost complete exclusion of genetic variation.

TABLE 3. VARIANCE

line	sternopleural	abdominal	degrees of freedom	ratio
high females	1.91042	6.57292	192	3.4405
high males	2.08163	6.75306	196	3.2441
low females	1.91346	6.85192	208	3.5809
low males	2.27451	5.80196	204	2.5508
pooled	2.04600	6.49300	800	3.1735

The test may, however, be made more precisely, as we know the responses to selection in each experiment. After seventeen generations the abdominals had changed in one line by rather more than one chaeta (an advance of just over 3%) but were unaltered in the other line. In a like period of the first experiment the sternopleurals also changed by just over one chaeta (here an advance of about 6%) in one line and were unaltered in the other. Thus the genetic variation had proved to be equally big, on an absolute scale, in the two cases, whereas the sternopleurals showed less than one-third the fluctuational variance of the abdominals. It therefore appears that fluctuation is relatively larger in the latter, as we should expect from the nature of the selective responses in the two experiments.

There are, of course, other agencies, such as dominance and interaction, which could help in sheltering variation, but segregation would lead to the exposure of a portion of the hereditary variation held by such agents, whereas fluctuation affects the selection of all genotypes to an extent dependent only on the relative magnitudes of the two types of variation, hereditary and non-hereditary. Hence these agents can hardly be considered so effective as fluctuation, whose efficiency in hiding variation will be increased by various methods of reducing recombination, such as exist in the male *Drosophila*. In this way any character which shows

non-heritable variation must, except in highly inbred small populations, always have a store of hereditary variation protected by these fluctuations from diminution by the action of selection.

5. CONCLUSIONS

The amount of heritable variation in a population is continually diminishing both by the action of selection and by random changes in the gene frequencies. There must, therefore, exist a counterbalancing agent tending to restore the variation lost in this way. Such an agent is to be found in mutation, by which new allelomorphs are continually arising. This is as true of polygenic as of any other type of genetic variation.

Tests of the rate of origin of hereditary variation by mutation were made on two characters, sternopleural and abdominal chaeta numbers, in highly inbred *Drosophila melanogaster*. Selection was made for increase and decrease in chaeta number in both positions. The number of sternopleural chaetae changed fairly smoothly with selection where changes were found, but the response to selection for abdominal number appeared in bursts separated by periods of stability.

Though attractive at first sight, the assumption that these bursts mark the occurrence of a few relatively major mutations is shown to be untenable. The periods of change must be attributed to recombination of new polygenic variations which had accumulated without giving rise to any selective advance. The comparative behaviour of the various high lines in the experiment on abdominal chaetae fully accords with this view.

Non-heritable fluctuation in chaeta number serves as the cloak under which small heritable variants accumulate. The response to selection occurs when these small variations are built up by recombination into combinations having a greater effect on the phenotype. There will thus be a minimum degree of heritable variability, which cannot be acted on or reduced by selection, and the level of this minimum will be determined by the magnitude of non-heritable fluctuation.

REFERENCES

- East, E. M. 1935 *Genetics*, **20**, 443-451.
Fisher, R. A. 1930 *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.
Fisher, R. A. 1939 *Ann. Eugen., Camb.*, **9**, 109-122.
Haldane, J. B. S. 1936 *J. Genet.*, **32**, 375-392.
Johannsen, W. 1909 *Elemente der exakten Erblichkeitslehre*. Jena: Fischer.
Lindstrom, E. W. 1941 *Genetics*, **26**, 387-397.
Mather, K. 1941 *J. Genet.*, **41**, 159-193.
Mather, K. 1942 *J. Genet.*, **43**, 309-336.
Wigan, L. G. 1942 Selection and back-selection in *Drosophila melanogaster* (in preparation).

The integument and moult cycle of *Tegenaria atrica* (Araneae)

By H. C. BROWNING, B.Sc., Ph.D.

Department of Zoology, University of Sheffield

(Communicated by V. B. Wigglesworth, F.R.S.—Received 6 March 1942)

The integument of *Tegenaria atrica* consists of two chitinous layers, an outer exocuticula and an inner endocuticula. The former is impregnated with protein and pigment and may be birefringent. Its thickness and development of birefringence is related to the hardness of each region.

The exocuticula is formed before, and the endocuticula after, the moult. The hypodermis appears to secrete the chitin of both layers. The substances impregnating the exocuticula are probably brought to it by granular blood cells. These granulocytes remain in the hypodermis after the moult.

The granulocytes responsible for the exocuticular secretion at the last moult migrate on to the old integument as this separates from the hypodermis at the next one. An ecdysial fluid is present and is absorbed by the time of the actual moult but very little digestion of the old integument occurs.

The chromatin of the hypodermal nuclei increases in amount up to the beginning of the secretion of the new integument. It then sharply decreases until after the moult, when the cycle is resumed.

There are three types of blood cell—granulocytes, leucocytes, and leberidocytes. The latter have a single large vacuole and are formed from the leucocytes. They only appear in relation to the moult, forming 65 % of the total blood cells immediately after it.

The digestive diverticula secrete a fluid which fills most of the alimentary canal at the time of the moult.

The above results are discussed. There is evidence that the relative hardness of the exocuticula is due to both impregnation with proteins and phenols and to changes of molecular structure. It is suggested that the leberidocytes absorb water from the food, via the plasma, to increase the blood pressure for moulting and subsequent enlargement of the animal. The filling of the alimentary canal with fluid is associated with this. The possible origin of a moulting hormone is considered.

1. INTRODUCTION

The integument and moult cycle have been described in the decapod Crustacea (*Homarus*) by Yonge (1932, 1936) and in the Insecta (*Rhodnius*) by Wigglesworth (1933, 1934a). I am indebted to Professor C. M. Yonge for suggesting that the problems of ecdysis should be investigated in the Araneae.

Wagner (1888) described the general biology of moulting in spiders and the histology of some organs. He states that the new integument is formed from the upper layers of the hypodermis which cease to stain and become chitinized, and that there is a moulting fluid which disappears several hours before ecdysis. Millot (1926) says that the hypodermal cells secrete the integument and that this consists of three layers. Schlottke (1938) found that the hypodermis has two types of nuclei, differing in their chromatin distribution but without apparent functional distinction. In the hypodermis of the lung vestibule only, he found glandular cells and considered that they form the moulting fluid, as their contents were emptied

at the moult. Wagner (1888), Kollmann (1908), Millot (1926), and Deevey (1941) have considered changes of the blood cells, and all agree that there is a special type of cell differentiated during moulting activity.

2. TECHNIQUE

The moult cycle was investigated in fifth-instar *Tegenaria atrica*, various points being confirmed from *T. durhamii*, *Lycosa lugubris*, *L. nigra*, and *Zilla litterata*. Except for general darkening a day or so before moult, spiders show no external changes during the cycle, and there is considerable variation in individual instar lengths (Bonnet 1930). This variation, and the effect of environmental factors on instar length, was investigated in *Tegenaria atrica* (Browning 1941). It was least at 30° C, 8.25 mm. saturation deficiency, and with a continual excess of food, when the average length of the fifth instar was 14.32 days. A number of *T. atrica* were kept under these conditions for this instar and several fixed each day. Experimentally, it was found that Susa at 30° C gave the best fixation, and most serial sections were stained according to Hollande's method (Langeron 1925).

At the same intervals preparations were made of the dorsal cephalothoracic, sternal, and abdominal hypodermis and blood samples taken. Smear preparations of the latter were unsatisfactory. A drop of blood was spread out on a cover-slip by needles and the slip placed over a glass ring cemented to a slide and containing a few drops of 1 % osmic acid. The blood cells were examined immediately and after haematoxylin-eosin staining. The various types were counted. Permanent preparations were made in glycerine jelly to avoid distortion due to shrinkage of the plasma.

3. THE INTEGUMENT

(a) Surface structure

The surface of the integument shows sinuous striae, about 1μ apart, which really are ridges of the outer layer (figures 1, 3A). They are much more conspicuous on the abdomen than on the cephalothorax, where they are difficult to detect in sections. On the former there are two types of ridge, every third one being thicker and 1μ in height, the other two about two-thirds of this. There are occasional anastomoses and the general course is displaced by the hair sockets (figure 1). Wagner (1888) described similar striations, stating that the pattern differed between genera and that there were rows of pores between the ridges. These could not be detected, but there were numerous lighter areas in this position on stained preparations (figure 2, l.a.).

The abdomen of spiders increases in volume to a considerable extent after feeding, but there are neither folds of the surface layer nor pleats of the entire wall to allow for this as in *Rhodnius* (Wigglesworth 1933). In *Zilla litterata* the striae were 2.9μ apart before feeding and 3.3μ after. The abdominal wall must therefore be extensible and actually does not feel rigid like that of the cephalothorax.

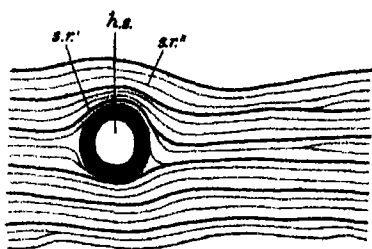


FIGURE 1. *T. atrica*, fifth instar. Surface view of abdominal integument. $\times 1000$.

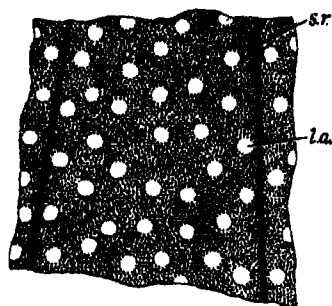


FIGURE 2. *Zilla litterata*, adult. Surface view of abdominal integument, hæmatoxylin-eosin. $\times 4625$.

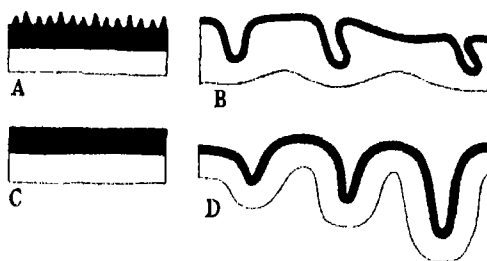


FIGURE 3. *T. atrica*, fifth instar. Transverse section of the integument of typical regions. $\times 1500$. A, abdomen; B, intersegmental membrane; C, cephalothorax; D, rostrum. (Exocuticula black, endocuticula white.)

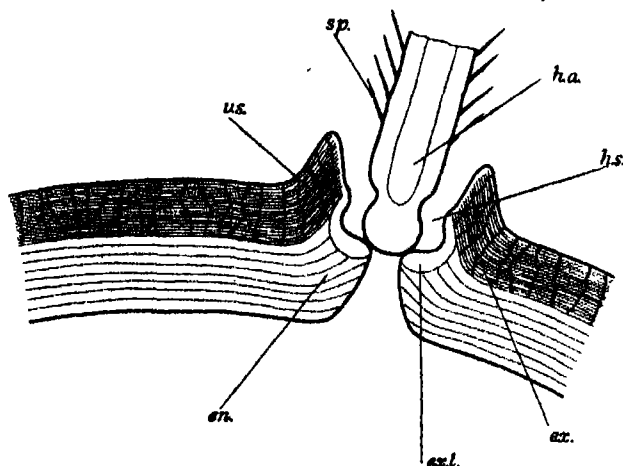


FIGURE 4

FIGURE 4. *T. atrica*, adult. Transverse section of integument and hair socket. $\times 680$.

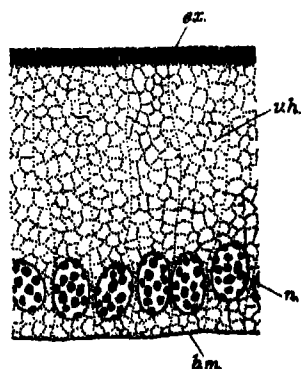


FIGURE 5

FIGURE 5. *T. atrica*, fifth instar. Transverse section of the hypodermis and integument of the anterior region of the maxillae. $\times 1185$.

b.m. basement membrane; en. endocuticula; ex. exocuticula; ex.l. unpigmented exocuticula lining hair socket; h.a. hollow axis of hair; h.s. hair socket; l.a. light area of stained exocuticula; n. hypodermal nucleus; sp. hair spine; s.r. surface ridge of exocuticula (s.r.') thick, s.r.'' thin; v.h. finely vacuolated hypodermis; v.s. vertical striation of exocuticula.

(b) *Internal structure*

In all regions, except the anterior surface of the maxillae, the integument consists of two layers, the outer basophil and the inner acidophil (figures 3, 4). Millot (1926), staining with haemalum, acid fuchsin, metanil yellow, and light green, found three layers and illustrates the outermost as about one-seventh of the total thickness and well defined. This layer could not be detected by any technique, including the above, but the staining reaction of the surface region of the outer layer often varied. For example, with Mallory, the inner layer is blue and the outer red with occasional gradations to yellow at the free edge; but the yellow parts are never sharply delimited. The outer layer has been called *exocuticula* and the inner *endocuticula*. The region of the maxillae already mentioned has *exocuticula* only (figure 5).

TABLE 1. THICKNESS (IN MICRA) OF EXOCUTICULA (*ex.*) AND ENDOCUTICULA (*en.*) IN THREE SPECIES OF SPIDERS

species	dorsal cephalothorax		ventral cephalothorax		abdomen	
	<i>ex.</i>	<i>en.</i>	<i>ex.</i>	<i>en.</i>	<i>ex.</i>	<i>en.</i>
<i>Tegenaria atrica</i> (fifth instar)	2	2.5	2	3	2	2
<i>Lycosa lugubris</i> (adult female)	6	4	3	5.5	5	5
<i>Zilla litterata</i> (adult female)	11	3	17.5	5	17	2.5

In *Tegenaria atrica* the two layers are roughly equal in thickness over most of the animal (table 1 and figures 3 A, C). But in the intersegmental membranes the *endocuticula* is thickened and the *exocuticula* reduced (figure 3 B). These membranes are found at the base of the spinnerets, form the petiole, and attach the appendages to the cephalothorax. They consist of 0.5μ *exocuticula* and $2.5\text{--}7\mu$ *endocuticula*, and the latter is thrown into folds which affect the surface of the membrane. This suggests that, relatively, the *exocuticula* is rigid and the *endocuticula* elastic. Actually the rostrum has a much reduced *exocuticula* (figure 3 D) and is soft unlike the other mouth-parts. Further, in regions requiring special firmness, the *exocuticula* is increased, for the fang of the chelicera has *endocuticula* 2μ and *exocuticula* 8μ thick, and spines and hairs are rigid hollow structures of the *exocuticula* only. On the other hand, the abdominal integument has both layers in equal proportions (table 1 and figure 3 A), but is extensible and comparatively soft.

The lens of the eyes is integumental with *exocuticula* 7μ and *endocuticula* 50μ thick (figure 6), the lining of the foregut shows the same layers with *exocuticula* innermost, and the endosternite consists of *endocuticula* only.

The relative proportions of the two layers differed between species (table 1). The greatly increased *exocuticula* of *Zilla litterata* may be due to its more exposed habitat and to the definitely cubical shape of the abdomen.

In both exocuticula and endocuticula there are horizontal striae, but these are more marked and farther apart in the latter (figure 4). Sinuous vertical striae run through the exocuticula (figure 4, *v.s.*) and resemble the 'pore canals' of *Rhodnius* (Wigglesworth 1933), where, however, they are confined to the inner layer of the integument. They did not fill with air when sections were dried and did not stain differently from the exocuticula. They apparently correspond to the light areas seen on the surface of this layer (figure 2). In the region of the maxillae with exocuticula only, much more conspicuous vertical striae, or pores, can be seen (figure 5), even in stained serial sections.

(c) *Microchemistry*

In unstained preparations the exocuticula is pale brown in general but colourless in the intersegmental membranes, the lens, the tips of the rostrum and labium, the special region of the maxillae, and the whole abdomen (except the spinnerets). The exocuticula of hairs is pigmented everywhere and the endocuticula is always colourless. The exocuticula is also birefringent in general with similar exceptions to those of pigmentation. However, the pigment is not responsible for birefringence, as the exocuticula of the lens, the tip of the labium, and the special region of the maxillae are birefringent although not pigmented. Small scattered patches of the abdominal exocuticula are also birefringent. Birefringence, as well as thickness of exocuticula, seems to be related to the rigidity of the part concerned.

The results of chemical tests (table 2) applied to pieces and sections of the integument show that both layers consist of chitin and that the exocuticula is impregnated with protein as well as its pigment. No differences were found between the integument of the cephalothorax and the abdomen. The same tests were carried out on the integument of *Rhodnius prolixus* for comparison.

The chitin test of Campbell (1929) was modified for treatment of thin sections. The material was heated in caustic potash (saturated solution at room temperature) for 15 min. at 150° C after which serial sections were cut in the usual way. After removal of the wax the sections were treated with iodine and sulphuric acid as the author describes. Both this and tests on large pieces showed that there was chitin in both layers, although the reaction of the exocuticula was less marked than that of the endocuticula when a weaker iodine solution was used. Tetragonal crystals of chitosan sulphate were obtained by the method given in the above paper. The iodine-calcium chloride and iodine-zinc chloride tests (Langeron 1925) did not give the reddish violet colours indicated.

In general, concentrated mineral acids dissolved the endocuticula in the cold and the exocuticula on heating. During the process the endocuticula swells considerably and the horizontal striations of both layers become more marked. A saturated solution of caustic potash did not dissolve the exocuticula, even after long heating, but rigidity was lost and the surface slightly attacked as the surface ridges became distorted and very faint. At the same time the pigment disappeared

and there were refringent droplets in the exocuticula which stained yellow with iodine but did not take up Sudan 3.

TABLE 2. TESTS APPLIED TO PIECES AND SECTIONS
OF THE INTEGUMENT OF *Tegenaria atrica*

test	result	
	exocuticula	endocuticula
CHITIN		
sat. KOH at 150° C for 15 min.—0.2 % iodine and 1 % H ₂ SO ₄	intense violet	intense violet
CaCl ₂ and iodine solution	red-brown	red-brown
Lugol's iodine and ZnCl ₂ solution	red-brown	red-brown
CELLULOSE		
Schweitzer's solution	nil	nil
MONOSACCHARIDE		
Fehling's solution	nil	nil
FAT		
1 % osmic acid solution	nil	nil
Sudan 3, alcoholic solution	nil	nil
HNO ₃ and KClO ₃ solution	dissolves on strongly heating	dissolves
—cerinic acid test		
dry heat	chars but does not fuse	chars but does not fuse
PROTEIN		
Millon's solution	intense red	nil or faint pink
xanthoproteic test	red-brown	nil or faint pink
ORTHO-DIHYDROXYPHENOL		
FeCl ₃ solution and NaHCO ₃ solution	nil	nil
ACTION OF ALKALI		
sat. KOH solution	pigment dissolves on heating, surface slightly attacked	nil
ACTION OF ACIDS		
H ₂ SO ₄ conc.	dissolves on gentle heating	dissolves without heating
HNO ₃ conc.	slowly dissolves on strongly heating	swells and slowly dissolves
HCl conc.	nil	swells and slowly dissolves on gentle heating

No reaction was obtained indicating the presence of fats, lipins, or waxes. Miranda (1904-5) found pores full of reduced copper in the integument after treatment with Fehling's reagent, but this was not confirmed. The iso-electric point of the endocuticula, determined by the method employed by Yonge (1932), was between 5 and 5.6, but the pigment of the exocuticula masked the reaction of this layer.

4. HYPODERMAL MOULT CYCLE

(a) General

The general features of the formation of the new integument are well seen in the lens where both exocuticula and endocuticula are very thick. In the resting period, from about the third day after moult to the tenth, the exocuticula is 7μ , the endocuticula 50μ , and the hypodermis 6μ in depth (figure 6A). Then the hypodermis deepens, reaching a maximum of 15μ during the 2 days before the moult (12th–14th days of the instar). Shortly after the beginning of this increase the old lens separates from the underlying hypodermis and eye cells, and secretion of the new exocuticula commences. By the fourteenth day this has reached the thickness of the old exocuticula (figure 6B), but it is concave at the lens and plicated elsewhere. It is acidophil and not basophil.

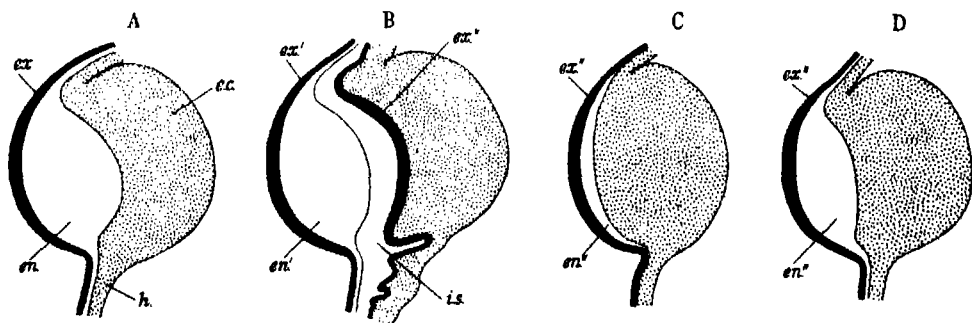


FIGURE 6. *T. atrica*, fifth instar. Longitudinal section of the lateral anterior eye. $\times 250$. A, resting stage—3rd–10th day of the instar; B, 12 hr. before the moult—14th day of the instar; C, 6 hr. after the moult; D, 2 days after the moult.

Shortly after the moult (figure 6C) the new exocuticula of the lens is convex and that of the rest of the integument smooth. Its reaction is now basophil, the hypodermis has decreased to 11μ , and a small amount of acidophil endocuticula has been secreted. By the second day after the moult the new endocuticula is 30μ thick at the lens and has been almost completely secreted elsewhere (figure 6D). The hypodermis is now 8μ thick. On the third day all the endocuticula has been formed and the hypodermis has returned to the resting condition (figure 6A).

(b) The resting hypodermis

The hypodermis (figures 7, 8A) is syncytial with finely granular cytoplasm and irregularly placed nuclei. It has an average depth of 6μ except in the rostrum where it is 15μ . There is a well-defined basement membrane separating it from the haemocoel and the cytoplasm is weakly acidophil. Pigment granules, from 0.3 to 2μ in diameter and yellow to purple-brown in colour, are present to some extent everywhere. They are usually placed near the integument (figure 8A, *p.*) and they, not the integumental pigment as stated by Savory (1928), give the colour pattern

of the animal (figure 7, *p.*). Large vacuoles (*v.*) occur irregularly in the cytoplasm. Only the hypodermis of the anterior surfaces of the maxillae and of the corresponding surfaces of the chelicerae differs from the rest. Here it is about 30μ in depth, has finely vacuolated cytoplasm and basal nuclei, no pigment, and is cellular (figure 5). Its position suggests that it is sensory.

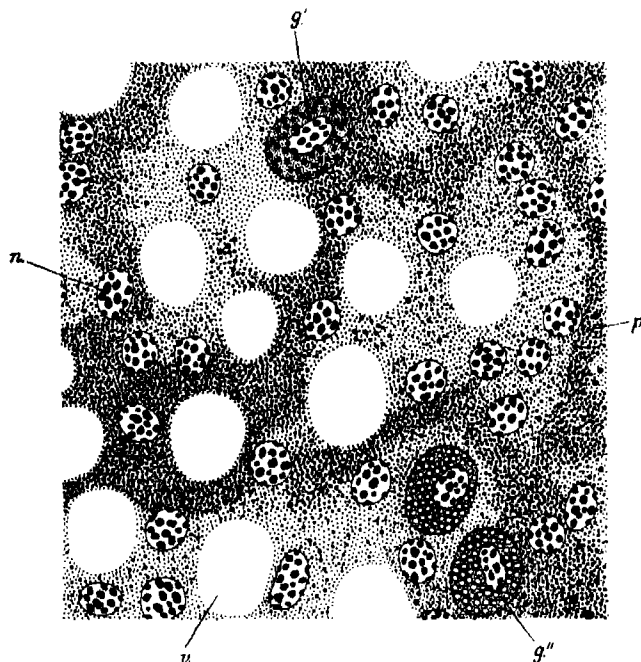


FIGURE 7. *T. atrica*, fifth instar. Sternal hypodermis, 3 days before the moult (11th day of the instar). $\times 1000$.

Scattered in all parts of the hypodermis are blood cells. These vary in size but are usually about 9 by 12μ in diameter (figures 7, 8A, *g'*). Their cytoplasm has numerous basophil granules (*b.g.*) of 0.5μ diameter and the nuclei are oval or kidney shaped, 6 by 4μ in dimensions. For reasons to be given later, these cells have been called 'old' granulocytes.

(c) Formation of exocuticula

Nine days after the previous moult 'new' granulocytes enter the hypodermis from the blood (figures 7, 8B, *g''*, *g''.h.*). They differ from the old blood cells in that their granules (*a.g.*) are slightly larger, refringent, acidophil, and more numerous, packing the cytoplasm. By the eleventh day they are at least as numerous as the old granulocytes, and the granules of the latter are now only weakly basophil, or even weakly acidophil. The hypodermis begins to deepen.

By the twelfth day the hypodermis has reached its maximum depth of 15μ (figure 8C). It becomes weakly basophil and has numerous small vacuoles of

basophil secretion (v'). It is separated from the old integument by fluid ($e.f.$) and a space ($i.s.$). New exocuticula begins to be secreted as a continuous acidophil layer (ex'') and, in suitable preparations, the surface ridges can be seen from the beginning. At the same time the new granulocytes (g'') have enlarged to about 17 by 13 μ and their nuclei to 7.5 by 5 μ . Their granules lose definition and in their places appear large vacuoles, full of acidophil secretion ($g''v'$).

Secretion of the new exocuticula continues until the next moult by which time it is fully formed (figure 8D). It is not birefringent nor pigmented. During this time the vacuoles in the new granulocytes become more numerous but smaller ($g''v'$), while their nuclei decrease in size to about 7 by 4 μ and the nucleoplasm becomes more chromatic. The small vacuoles of basophil secretion (v') in the hypodermis also increase in number, especially near the new exocuticula.

(d) *The old integument*

The inner surface of the old endocuticula (figure 8C, D, en') is covered by a thin layer of ecdysial fluid ($e.f.$) which contains small vacuoles or hyaline inclusions ($e'f'$). This fluid diminishes in amount, and its inclusions in number towards the moult and an emerging spider is quite dry. There is also a space between the old and new integument ($i.s.$), but possibly this is an artefact, as there does not appear to be air between the integuments in the living animal.

As the integument becomes free the old granulocytes pass from the hypodermis on to it and are found in the ecdysial fluid (figure 8C, D, g'). Their granules are gradually replaced by acidophil vacuoles ($a.v.$) of slightly larger size and the nucleoplasm becomes more chromatic. Degenerative changes are visible in some of the cells and their nuclei.

During these processes the inner edge of the endocuticula ($en'e.$) becomes uneven and small splits ($en's.$) appear in it. This, and the fact that the old endocuticula of the lens is appreciably thinner just before moult (figure 6B, en'), suggests that there is slight digestion of the old endocuticula by the fluid.

(e) *Formation of endocuticula*

A few hours after the moult (figure 8E) the exocuticula has become basophil and birefringent and a very thin layer of acidophil endocuticula (en'') has been secreted. This change of the exocuticula is not due to contact with the air as exocuticular hairs and hair sockets often show this staining reaction and birefringence before the moult. The integument as a whole is still soft but hardens during the first 12 hr. The hypodermis has decreased to 10 μ , is only very weakly basophil, and has very few secretory vacuoles (v'). The new granulocytes are smaller (14 by 10 μ), and their cytoplasm has only a few empty vacuoles ($g''v.$) and some small granules ($r.g.$). Their nuclei have returned to normal.

By the second day (figure 8F) almost all the endocuticula has been secreted and the hypodermis is 8.5 μ in depth, weakly acidophil, and has no vacuoles of secretion. The blood cells (g'') have no vacuoles but numerous small granules ($r.g.$) and have

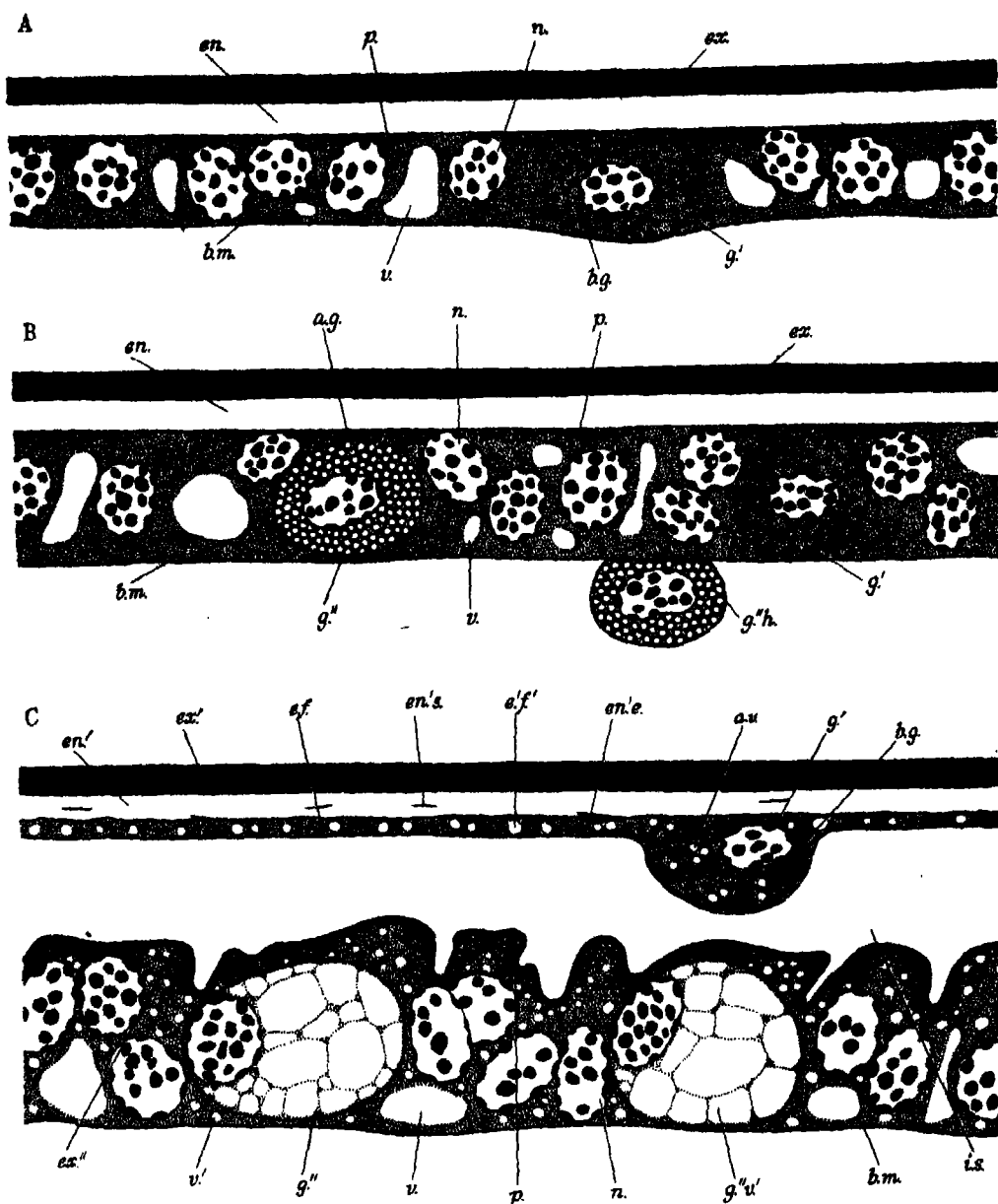


FIGURE 8. *T. atrica*, fifth instar. Transverse section of the integument and hypodermis. $\times 1750$.

A, resting stago—3rd–10th day of the instar; B, 3 days before the moult—11th day of the instar; C, 2 days before the moult—12th day of the instar; D, 12 hr. before the moult—14th day of the instar; E, 6 hr. after the moult; F, 2 days after the moult. *a.g.* acidophil granule of new granulocyte; *a.v.* acidophil vacuole of old granulocyte; *b.g.* basophil granule of old granulocyte; *b.m.* basement membrane; *e.f.* ecdysial fluid (*e.f'.* inclusion); *e.c.* eyecells;

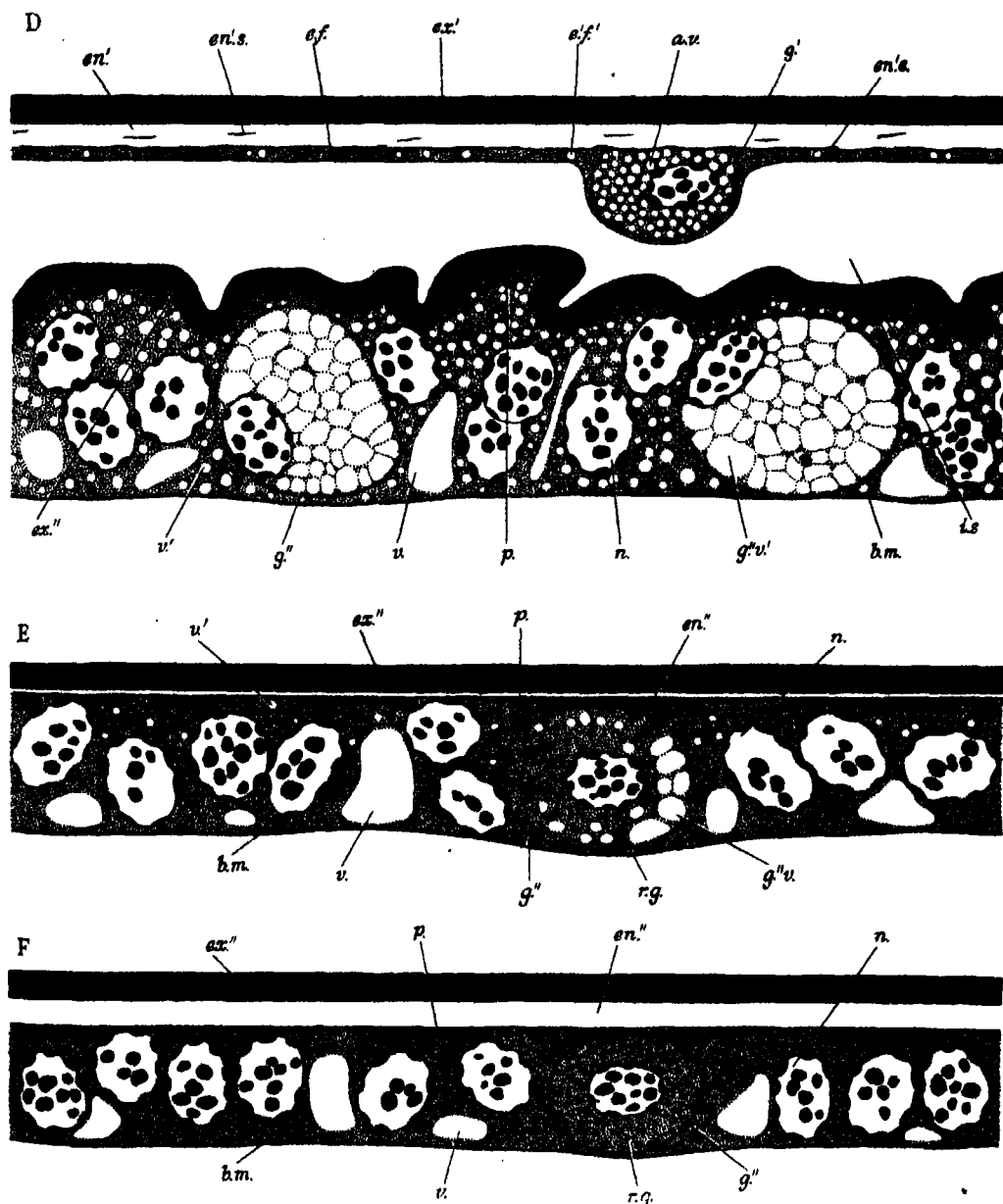


FIGURE 8 (continued)

en. endocuticula (*en'*. old, *en''*. new); *en'.e.* uneven inner edge of old endocuticula; *en'.s.* split in old endocuticula; *ex.* exocuticula (*ex'*. old, *ex''*. new); *g.* granulocyte (*g'*. old, *g''*. new); *g''.h.* new granulocyte entering hypodermis from haemocoel; *g''.v.* vacuole in new granulocyte (*g''.v.* empty, *g''.v'*. full of secretion); *h.* hypodermis; *i.s.* space between new and old integuments; *n.* hypodermal nucleus; *p.* pigment granules; *r.g.* reappearing granules in new granulocyte; *v.* large hypodermal vacuole; *v'*. small hypodermal vacuole full of secretion.

decreased to the pre-moult size. By the third day all the endocuticula has been formed, the hypodermis is in the resting stage, and the granules of the blood cells are basophil (figure 8A). These cells are now the 'old' granulocytes of the resting hypodermis. The pigment of the exocuticula develops during the first 3 days after the moult and has now reached its maximum depth.

Schlottke (1938) states that there are glandular cells in the hypodermis of the lung vestibule only. From his description and figures these are evidently granulocytes, and their scattered occurrence (figure 7) account for the fact that this author did not find them in other regions.

(f) *Nuclear changes*

In the resting state (figure 8A) the nuclei of the hypodermis are about 6 by 4 μ , usually oval but sometimes round. They have a well-defined nuclear membrane and a number of irregular chromatin masses, of an average diameter of 1 μ . As stated by Schlottke (1938) two types of nuclei can be distinguished, one with well-defined chromatin, the other slightly smaller but with ill-defined chromatin and

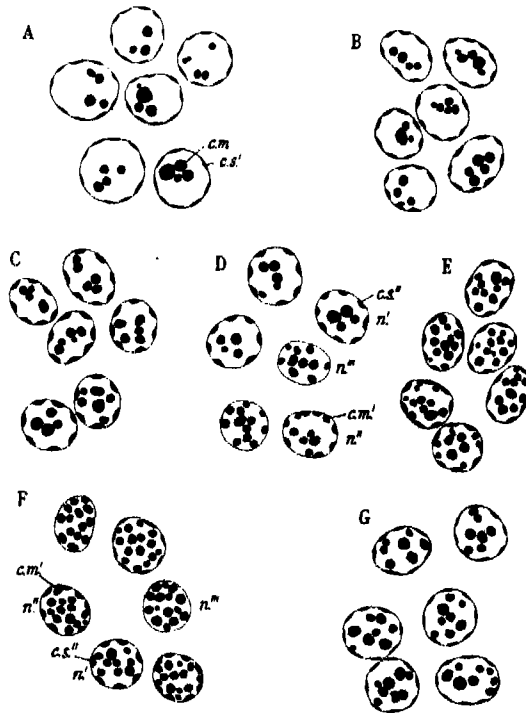


FIGURE 9. *T. atrica*, fifth instar. Cycle of chromatin changes in the hypodermal nuclei. $\times 1200$. A, 1st day; B, 2nd day; C, 3rd-4th day; D, 5th-6th day; E, 7th-10th day; F, 11th-12th day; G, 13th-14th day. c.m. chromatin mass; c.m'. chromatin mass, just formed; c.s'. chromatin swelling of nuclear membrane, early; c.s''. chromatin swelling of nuclear membrane, late; n'. hypodermal nucleus about to increase the number of chromatin masses; n''. hypodermal nucleus increasing the number of chromatin masses; n'''. hypodermal nucleus immediately after increase of the number of chromatin masses.

more chromatic nucleoplasm. These differences appear to have no significance and all intermediary stages can be seen.

During moulting activity changes of nuclear size lag slightly behind those of the hypodermis. The nuclei do not enlarge with the initial deepening of the latter but, by the time a little new exocuticula has been secreted, they have grown to about 7 by 5μ (figure 8C). They remain thus throughout the formation of the exocuticula and of most of the endocuticula (figures 8D, E). By the second day of the new instar, when almost all the endocuticula has been secreted, they have decreased to about 6.5 by 4.5μ (figure 8F). On the third day they are normal.

During the instar there is a definite cycle of increase and decrease of nuclear chromatin, giving a maximum immediately before the secretion of new exocuticula and a minimum just after the moult. On the first day of the instar, each nucleus has three or four chromatin masses (figure 9A, *c.m.*), and the nuclear membrane has a number of slight chromatic swellings (*c.s'*). From the second to the fourth day the number of chromatin masses increases by two or three and the swellings of the membrane enlarge (figure 9B, C). By the fifth or sixth days the peripheral swellings (figure 9D, *c.s''*) have reached the size of the free masses. They begin to be detached (*c.m'*) to form new chromatin masses, and the nuclear membrane, temporarily, has no swellings. Nuclei in all stages of the process can be seen at this time (figure 9D, *n'*, *n''*, *n'''*). Consequently, by the seventh day (figure 9E), all the nuclei have from nine to twelve chromatin masses. During the next 5 days the process is repeated so that by the twelfth day each nucleus has from nine to fifteen masses (figure 9F). From this time to the moult new exocuticula is being secreted, and the number of free masses decreases to that found immediately after the moult (figure 9G).

Occasional mitoses were seen throughout the instar, but there was no simultaneous mitotic activity of even a number of nuclei at any time.

5. BLOOD-CELL MOULT CYCLE

(a) *Types of blood cell*

There are three main types of blood cell—granulocytes, leucocytes, and leberidocytes. In the resting stage there are about 28 % of the first type and 72 % of the second. Leberidocytes are only found during moulting activity. Figures 10 and 11 show twenty representative cells in the resting stage and immediately after the moult.

The granulocytes (*g.*) are round or oval, vary in diameter from 10 to 20μ , and usually have an oval or kidney-shaped nucleus. The cytoplasm may emit needle-like pseudopodia and, apart from these, is packed with refringent uniform granules of 0.5μ diameter. In any blood sample the granules of some cells take up eosin, others stain from light to dark brown with osmic acid, and some do not stain at all. The granulocytes are formed from small leucocytes which may be seen with a few to many of the typical granules.

The leucocytes (*l*, *l'*.) are round or oval with a round nucleus and acidophil cytoplasm. They vary in size from those consisting of nucleus and a very thin layer of cytoplasm to those 20μ in diameter. Some of the largest differ in that their cytoplasm is basophil and markedly granular. They may emit short club-shaped pseudopodia. There are three subtypes. The first (*l*.) has no inclusions in the cytoplasm and forms two-thirds of the total number. The second has two or three inclusions in the cytoplasm; these are refringent, osmophil, and about 2μ in diameter, although showing much variation in size. This subtype only occurs occasionally and then in very small numbers. The third (*l'*.) has hyaline non-staining

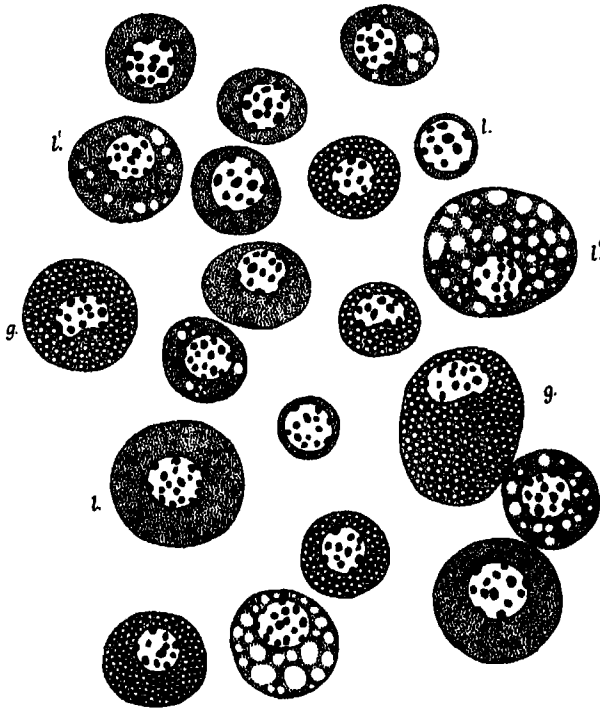


FIGURE 10. *T. atrica*, fifth instar. Twenty representative blood cells, 8th day of the instar. $\times 1000$.

inclusions or vacuoles varying from 0.5 to 3μ in diameter, and such cells may have anything from a few to numerous inclusions. About a third of the leucocytes belong to this subtype. In general, inclusions tend to occur mainly in the larger leucocytes.

The leberidocytes (figure 11) develop from leucocytes of about 12μ diameter which have no inclusions. Their cytoplasm becomes very hyaline, and a number of refringent granules of varying size (0.25 – 1.5μ) appear in it (*lb'*). These do not take up any stains but are always slightly osmophil. One, or sometimes two, small vacuoles, without staining contents, appear and gradually increase in size (*lb''*, *lb'''*). Finally, the cell consists of one large vacuole surrounded by a thin layer of

cytoplasm with the nucleus, oval and flattened, occupying one pole (*lb.*). The cell has increased to 25 or 30 μ in diameter and the nucleus from 7 by 6 μ to 10 by 6 μ . The inclusions are unaltered throughout. The vacuole contains fluid and, when the differentiated cells are mounted in water, it enlarges further and the cell bursts. The leberidocytes regress to leucocytes by the reversal of the above processes.

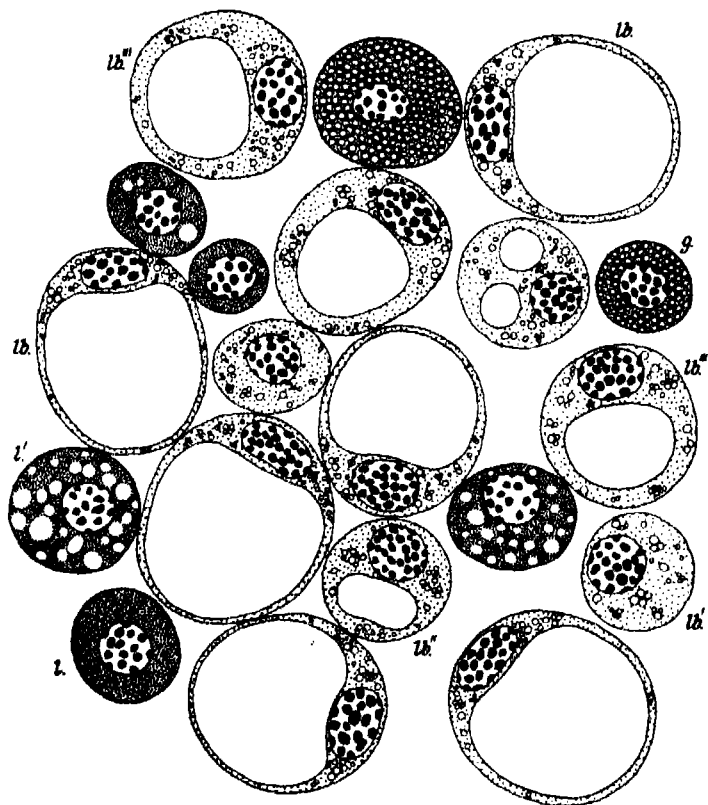


FIGURE 11. *T. atrica*, fifth instar. Twenty representative blood cells, 1st day of the instar (6 hr. after the moult). $\times 1000$. *g.* granulocyte; *l.* leucocyte; *l'* leucocyte with hyaline inclusions; *lb.* fully differentiated leberidocyte; *lb'* leberidocyte, earliest stage with very hyaline cytoplasm but no vacuole; *lb''* leberidocyte with small vacuole; *lb'''* leberidocyte with medium vacuole.

(b) Variation of the cell formula

Blood samples were taken daily through the fifth instar and the various types of cell counted (table 3). The granulocytes form only 10 % of the total immediately after the moult but, by the eighth day, they have increased to 32 %. Probably most of this increase is due to differentiation from leucocytes but, as the hypodermal blood cells are never as numerous just after the moult as just before it (figure 8), some may re-enter the haemocoel in the first few days of the instar. From the tenth day onwards new granulocytes are entering the hypodermis and

the free granulocytes have decreased to 22 % by the twelfth day when secretion of exocuticula commences. The figures suggest that they continue to enter the hypodermis during the latter process.

TABLE 3. BLOOD FORMULA OF *Tegenaria atrica* DURING THE FIFTH INSTAR

day of instar	leuco- cytes %	granulo- cytes %	leberido- cytes %	day of instar	leuco- cytes %	granulo- cytes %	leberido- cytes %
first	27	10	63	eighth	68	32	0
second	52	17	31	tenth	67	30	3
third	65	21	14	twelfth	60	22	18
fourth	79	20	1	thirteenth	48	23	29
sixth	75	25	0	fourteenth	41	19	40

The leberidocytes are not present at all from the fifth to the ninth day, and there is only a very small percentage on the fourth and tenth days, but evidently their period of activity is slightly longer than that of the hypodermis. The percentage increases to 40 % just before the moult (fourteenth day) and to 63 % just *after* the moult. It then steadily decreases until all the leberidocytes have regressed to leucocytes.

The leucocyte percentage varies in accordance with the origin and regression of the leberidocytes and differentiation of granulocytes.

Figures 12 and 13 are drawings of actual sagittal sections of the cephalothorax of *T. atrica*, during the resting stage and just after the moult respectively. Any differences other than those of the blood cells or midgut caecum are due to slight differences of plane. In the resting stage the blood spaces (*b.s.*) have cells either in small groups or widely separated, but in the post-moult stage all blood spaces are packed with cells, mainly leberidocytes. This is due to the change of cell formula, for the greater number (299) of cells in figure 13 than in figure 12 (188) is easily accounted for by the fact that a section will pass through more cell walls when the cells are larger (if the diameter of a cell is doubled the volume will be increased by eight times). Further, while occasional mitoses were seen in the blood cells during the instar, no general mitotic activity took place.

(c) Blood cells in other spiders

The blood cells and changes of the formula have been described in *Tarantula* (Wagner 1888), in *Tegenaria atrica* (Kollmann 1908), in *Amaurobius similis* and other species (Millot 1926), and in *Phormictopus cancerides* (Deévey 1941). The three main types of cell were found by all authors. Their synonyms and special types found in addition are listed in table 4.

In regard to the special types only Wagner mentions 'balloons' which he does not describe in detail. Plasma artefacts occur which would justify this description. Kollmann considered that the nephrocytes of the cephalothorax (see figures 12, 13) became free and formed leberidocytes at the time of moulting, calling them

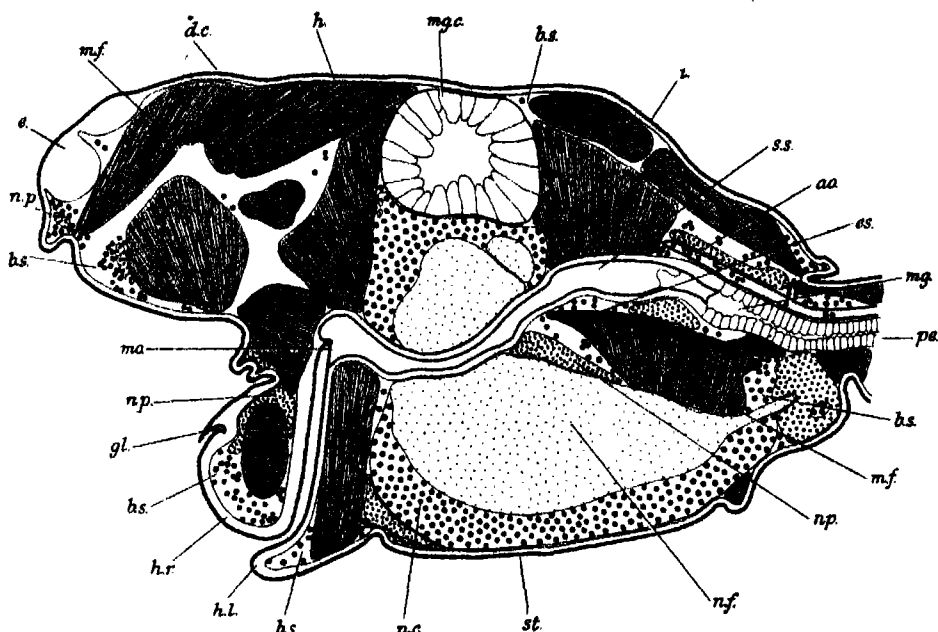


FIGURE 12. *T. atrica*, fifth instar. Sagittal section of the cephalothorax in the resting stage (3rd-10th day of the instar.) $\times 62.5$.

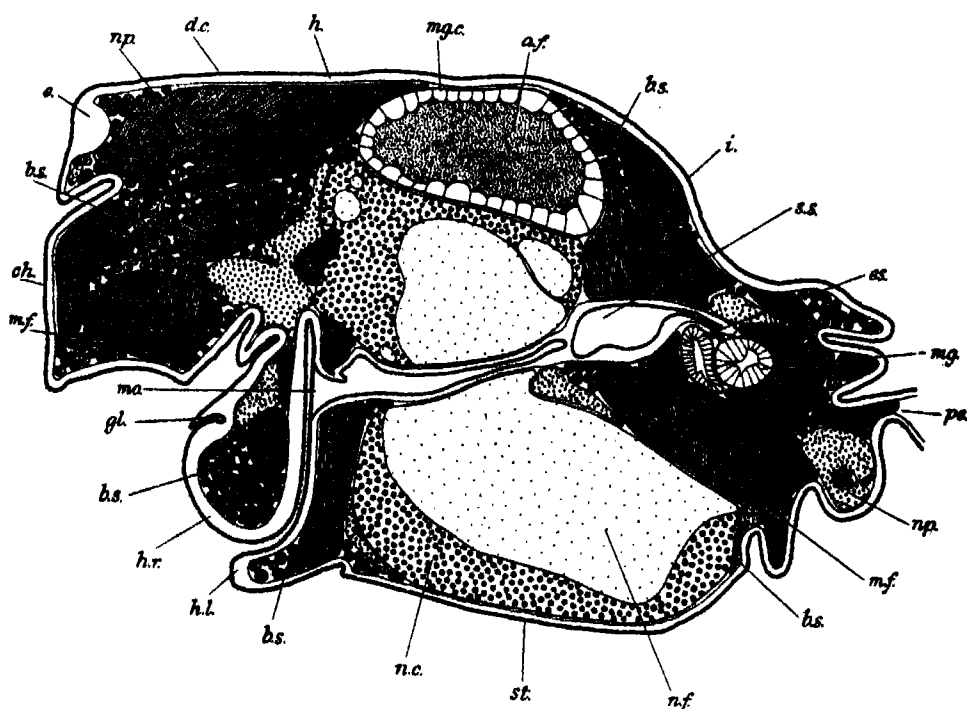


FIGURE 13. *T. atrica*, fifth instar. Sagittal section of the cephalothorax 6 hr. after the moult. $\times 62.5$.

af. acidophil fluid in midgut caecum; *ao*. aorta; *bs*. blood space; *ch*. chelicera; *dc*. dorsal cephalothorax; *e*. eye; *es*. endosternite; *gl*. rostral gland; *h*. hypodermis; *h.l*. deep hypodermis of labium; *h.r*. deep hypodermis of rostrum; *i*. integument; *mf*. muscle fibres; *my*. midgut; *mg.c*. midgut caecum; *mo*. mouth; *nc*. nerve cells; *nf*. nerve fibres; *np*. nephrocytes; *pe*. petiole; *ss*. sucking stomach; *st*. sternum.

'supplementary cells'. However, the two types of cell differ considerably, there are no transition stages, and there is no mitotic activity of the nephrocytes during an instar, although they are far less numerous than the leberidocytes. Millot, dealing with species other than *Tegenaria atrica*, also disagreed with Kollmann for similar reasons. Cuénot (1891, 1897) and Millot found cells with protein crystals, but the latter author states that they were never numerous and that they only occur in some species.

TABLE 4. SYNONYMS AND SPECIAL TYPES OF CELLS FOUND IN ARANEID BLOOD

author	granulocytes	leucocytes	leberidocytes	special types
Wagner	coloured cells	amoeboid cells	spherical cells with vacuole	balloons
Kollmann	granular leucocytes	hyaline leucocytes	vacuolar cells	supplementary cells
Millot	granulocytes, granular leucocytes	hyaline leucocytes (if very small —lymphocytes)	vacuolocytes, vacuolar cells	cells of Cuénot
Deevey	chromophobes, eosinophils, and weak eosinophils	hyaline leucocytes	leberidocytes	basophils and 'cyst' cells

Deevey found cells containing basophil crescent-shaped granules and 'cyst' cells which appear to be peculiar to *Phormictopus*. She divides the granulocytes into chromophobes, eosinophils, and weak eosinophils, the names referring to the inclusions. As this author states that there may be more than one type of granule in the same cell, and as all three types occur in any blood sample, they appear to be stages in the differentiation of one type only. Deevey proposed the name 'leberidocyte' for the type of cell only appearing in relation to moulting, as such cells in *Phormictopus* do not have a vacuole but numerous chromophobic inclusions and a few refringent granules.

All authors agree that the leberidocytes greatly increase at the time of the moult. Wagner and Kollmann considered that some leberidocytes were always present in the blood, and the latter that these were unimportant, the increase being due to the nephrocytes. This was probably due to lack of accurate information of the position of the spider in the moult cycle. Millot found that there were none in the resting stage, and that they formed 90 % of the total from the time of the moult to 20 hr. after it. This very high percentage may have been an injury effect, as successive blood samples were taken from the same animal. Deevey did not make determinations at very close intervals of time, but her figures indicate that the highest percentage (47 %) is reached at the moult and sustained for a short time after it. She found no leberidocytes in the resting stage, and noticed that the granulocytes increased during this time and decreased during the differentiation of the leberidocytes although, as the author points out, certainly not forming them.

6. ALIMENTARY CANAL AND MOULT CYCLE

Tegenaria atrica will eat four *Drosophila melanogaster* each day for the first 10 days of the fifth instar (Browning 1941). The abdomen enlarges considerably. From this time until the second day after moulting the spider will not feed and, during the last 4 days of the instar, the abdomen decreases in volume to some extent.

The essential parts of the alimentary canal are the ectodermal sucking stomach (figures 12, 13, *s.s.*) leading into the midgut (*mg.*) from which a large number of digestive diverticula arise. A pair of these, which joint anteriorly (*mg.c.*) and branch, are found in the cephalothorax, although the majority occupy the abdomen. During the last 4 days before the moult (10th–14th days of the instar), all food in the lumen of the diverticula is absorbed by the cells lining them. These cells then secrete an acidophil fluid (figure 13, *a.f.*) which fills all the diverticula, and sometimes the stomach and midgut itself, in the first few hours after the moult. It is at this time that the leberidocytes reach their maximum development and all blood spaces are packed with them (figure 13). This fluid is resorbed in the following 24 hr. and the animal will once more feed.

7. DISCUSSION

The integument consists of two layers, exocuticula and endocuticula, the former being secreted before the moult and the latter after it. Both contain chitin, but the exocuticula is usually pigmented and birefringent and is impregnated with protein. The thickness of this layer and its molecular arrangement are responsible for the varying hardness of the different regions of the integument.

The chitin is secreted by the hypodermis as the latter shows activity throughout the formation of both layers. Blood cells—the granulocytes—enter the hypodermis and show secretory activity during the formation of the exocuticula and, evidently, bring the impregnating substances. However, the pigment does not come from this source as it is not developed until after the moult. These cells cannot, as Schlottke (1938) supposes, secrete ecdysial fluid, as they are in full activity while this is disappearing.

The granulocytes, for the most part, remain in the hypodermis after their activity has finished and reform granules which are basophil, not acidophil. At the next moult they migrate on to the old integument, when this is freed from the hypodermis, and a new generation of granulocytes takes their place. Very little digestion of the old integument occurs, but there is an ecdysial fluid. This appears to be secreted by the hypodermis as, although the granules of the blood cells on the old integument are replaced by vacuoles, this change takes place after the fluid has appeared.

Pryor (1940 *a, b*) states that a dihydroxyphenol and a protein are secreted into insect integument and are responsible for its hardening. Millot (1926) thought that the granules of the granulocytes were albuminoid and were capable of association

with fats and lipins; and Deevey (1941) found some indications of phenols in the eosinophil granulocytes, i.e. the granulocytes which enter the hypodermis. Millot & Jonnart (1933) also found free 'phenol bodies' in the blood in connexion with moulting. Further, Bonnet (1930) found that, in wound regeneration, granulocytes accumulate at the site of injury and are transformed into integument. He does not state if the two normal layers are differentiated.

On the other hand, Fraenkel & Rudall (1940) have explained the hardening of the integument in the pupation of *Sarcophaga falcitata* (Diptera) as being due to the orientation of chitin crystallites and their binding by impregnating proteins. It is suggested that the process of hardening of the exocuticle of *Tegenaria atrica* is similar, for definite, as opposed to random, molecular arrangement takes place in this layer after moult, giving birefringence. This is not conditional upon exposure to air, as hairs and their sockets often show the change before moult, and a change of impregnating proteins would explain the altered reaction to stains after moult. Regions of the integument that remain soft show no molecular arrangement and the latter is independent of pigmentation.

The two layers of the integument resemble the impregnated and non-impregnated endocuticle of *Rhodnius* (Wigglesworth 1933). Here variation in hardness is accomplished by the presence or absence of impregnation rather than by relative thickness, but there is no evidence regarding molecular arrangement. There is no layer corresponding chemically to the cuticle of *Homarus* (Yonge 1932) or to the epicuticle of *Rhodnius* but, presumably, the exocuticle serves the functions of these, at least in regard to permeability or desiccation. The only point of similarity between *Tegenaria* and *Homarus* is in the migration of the 'old' granulocytes on to the old integument, for Yonge found nuclei concerned in the digestion of this in the crustacean. He did not investigate their origin. In *Tegenaria* their main function may be excretion. In *Limulus*, the only other arachnid that has been investigated (Yonge, unpublished), the integument is similar to that of *Homarus*. This and the resemblance of *Tegenaria* to *Rhodnius* may be examples of convergent evolution in different groups occupying the same habitat, while having the same basic structure to limit their possible adaptations.

Wagner (1888) considered that the leberidocytes were a stage in the proliferation of the other blood cells, stating that after the moult they divide and form granulocytes and leucocytes. Apart from the extreme improbability of a highly differentiated cell giving rise to a less differentiated type, such divisions were never seen. Deevey (1941) suggests that the leberidocytes are concerned in chitin secretion as they contain glycogen, and there is some evidence that this substance is one of the raw materials in its formation (Paul & Sharpe 1919; Needham 1929, etc.). It seems unlikely that these cells could be transporting glycogen to the hypodermis, as most of them are not in contact with it (figure 13). But Deevey also found glycogen in all the cell types at moult, although to a lesser degree than in the leberidocytes, and it seems more probable that glycogen is present in the plasma itself and is thus reaching the hypodermis.

An arthropod must have some means of increasing its internal pressure for the process of moulting and for growth before the new integument hardens. Insects usually use either air or fluid in the gut. For instance, *Cimex* shifts the intestinal contents from place to place and takes in bubbles of air (Kemper 1931); the sole function of the gut of adult Ephemeroptera and many Lepidoptera is to take in air for moulting (Wigglesworth 1934*b*); aquatic insects swallow water, and *Calliphora* expands on emergence by air intake (Wigglesworth 1939). In decapod Crustacea there is considerable absorption of water at the moult (Paul & Sharpe 1919; Robertson 1937), apparently for the same purpose. These methods are not available to spiders, for, at least under experimental conditions, they do not drink, and the gut has no chitin-lined extensible crop, food being stored in the cellular digestive diverticula. It is suggested that the leberidocytes develop by absorbing water from the plasma. This water originates in the food taken into the alimentary canal in the period before the moult, and by this means the total volume of blood is increased, the plasma remaining physiologically constant. This increased volume of blood will facilitate pressure effects by muscular contractions of the body wall, enabling these effects to be applied regionally for moulting. In this way would be invoked a mechanical principle somewhat similar to that operating in insects where, however, increased blood pressure can be obtained by inflation of part of the alimentary canal, whose structure permits this. This view is supported by the fact that the maximum development of leberidocytes is just *after* the moult, and that instar length and the amount of food required to complete an instar are both increased at low humidities (Browning 1941).

Wagner (1888) suggested that there must be an agent causing the general periodic changes of moulting. The biology of an instar of *Tegenaria atrica* indicates the release of a hormone dependent on the dilatation of the abdomen by food (Browning 1941). The time of its release would be before the differentiation of any leberidocytes. Bruntz (1904) and Millot (1930) found that the nephrocytes consist of two types of cell, one binucleate and about 40μ in diameter, the other uninucleate, much smaller, and with denser cytoplasm. Millot (1930) found no evidence that the latter were excretory and called them 'endocrine cells'. But neither type shows any rhythmic activity in relation to moult, as Millot admits, and it appears more likely that the so-called endocrine cells are young stages of the nephrocytes proper, especially as intermediary forms can be seen. Millot (1926) found no changes of the pH of the plasma corresponding to leberidocyte differentiation. Experiments on the injection of the blood of moulting spiders into non-moulting animals, both by Millot (1926) and Deevey (1941), gave negative results.

There are several sympathetic ganglia in the cephalothorax of spiders whose function is unknown (Gerhardt & Kästner 1937). The stomodaeal bridge in particular resembles the corpora allata of insects, which have been shown to secrete hormones (Wigglesworth 1934*a*; Burt 1937). However, no histological evidence of secretion could be obtained and, in any case, this would not be conclusive without experimental evidence. The delicacy of spiders makes it difficult, if not

impossible, to perform suitable grafting experiments, and the great variation in instar length (Bonnet 1930; Browning 1941) may be responsible for the lack of positive results in the experiments of Millot and Deevey, especially in the small numbers of animals that they used.

REFERENCES

- Bonnet, P. 1930 *Thès. Fac. Sci. Toulouse*, **44**, 1.
 Browning, H. C. 1941 *Proc. Zool. Soc. Lond. A*, **111**, 303.
 Bruntz, L. 1904 *Arch. Biol., Paris*, **20**, 216.
 Burt, R. T. 1937 *Proc. Roy. Soc. B*, **124**, 13.
 Campbell, F. L. 1929 *Ann. Ent. Soc. Amer.* **22**, 401.
 Cuénot, L. 1891 *Arch. Zool. exp. gén.* (2) **9**, 13.
 Cuénot, L. 1897 *Arch. Anat. micr.* **1**, 153.
 Deevey, G. B. 1941 *J. Morph.* **68**, 457.
 Fraenkel, G. & Rudall, K. M. 1940 *Proc. Roy. Soc. B*, **129**, 1.
 Gerhardt, U. & Kästner, A. 1937 Kükenthal and Krumbach, *Handbuch der Zoologie*, Lief. 11. Berlin and Leipzig: Gruyter.
 Kemper, H. 1931 *Z. Morph. Ökol. Tiere*, **22**, 53.
 Kollmann, M. 1908 *Ann. Sci. Nat.* (9) **8**, 1.
 Langeron, M. 1925 *Précis de microscopie*. Paris: Masson.
 Millot, J. 1926 *Bull. Biol. Fra. Belg.* suppl. **8**, 1.
 Millot, J. 1930 *Arch. Anat. micr.* **26**, 43.
 Millot, J. & Jonnart, R. 1933 *C.R. Acad. Sci., Paris*, **197**, 1002.
 Mirande, M. 1904-5 *Arch. Anat. micr.* **7**, 207.
 Needham, D. M. 1929 *Biol. Rev.* **4**, 307.
 Paul, J. H. & Sharpe, J. S. 1919 *Biochem. J.* **13**, 487.
 Pryor, M. G. M. 1940a *Proc. Roy. Soc. B*, **128**, 378.
 Pryor, M. G. M. 1940b *Proc. Roy. Soc. B*, **128**, 393.
 Robertson, J. D. 1937 *Proc. Roy. Soc. B*, **124**, 162.
 Savory, T. H. 1928 *The biology of spiders*. London: Sidgwick and Jackson.
 Schlottke, E. 1938 *Z. Morph. Ökol. Tiere*, **34**, 207.
 Wagner, W. 1888 La mue des Araignées. *Ann. Sci. Nat.* (7), **6**, 281.
 Wigglesworth, V. B. 1933 *Quart. J. Micr. Sci.* **76**, 269.
 Wigglesworth, V. B. 1934a *Quart. J. Micr. Sci.* **77**, 191.
 Wigglesworth, V. B. 1934b *Insect physiology*. London: Methuen.
 Wigglesworth, V. B. 1939 *The principles of insect physiology*. London: Methuen.
 Yonge, C. M. 1932 *Proc. Roy. Soc. B*, **111**, 298.
 Yonge, C. M. 1936 *Proc. Roy. Soc. B*, **120**, 15.

The development of the bristles in normal and some mutant types of *Drosophila melanogaster*

BY A. D. LEES AND C. H. WADDINGTON

Zoological and Strangeways Laboratories, Cambridge

(Communicated by J. Gray, F.R.S.—Received 23 March 1942)

This paper describes the development of the normal macro- and micro-chætae of *Drosophila*, together with that of twelve mutant types. The phenotypes of twenty combinations of these genes have been studied.

Each normal bristle is secreted by a single cell, the trichogen, which lies beneath a tormogen cell which secretes a socket. These bristle cells are first distinguishable in the epidermis at about 15 hr. after puparium formation, when they have already divided to form a pair, and are slightly larger than the normal epidermal cells. The secretion of the bristle proceeds most rapidly between 30 and 55 hr., during which time the bristle cells are very large and obviously highly polyploid. The socket, apparently, does not completely enclose the base of the bristle in the earliest stages. The development of the microchætae is essentially similar to that of the macrochætae.

The actions of the twelve genes can be summarized as follows:

Scute causes a primary absence of certain bristle cells, and *extra-bristle-complex-41c* and *hairy* the presence of supernumerary groups. *Split* frequently causes an extra division, so that a group of four cells is formed; these may be arranged as two trichogens and two tormogens, or one trichogen and three tormogens; or the whole group may fail to reach the surface of the epithelium, when no bristle or socket is formed. *Dichaete* may produce an effect similar to the last-described of *split*, and it may also cause an extra division of the trichogen, producing a double bristle in a single socket. *Hairless* causes the trichogens of some bristle groups to lie level with the tormogens, and to develop like them into sockets. In *Stubble* the tormogens are shifted rather to one side of the trichogens, so that the bristle is less closely invested by the socket, and becomes thicker and shorter. In *shaven-naked* the trichogen is irregularly displaced, becoming more or less converted into a tormogen; the small bristle which may be secreted is often peculiarly fanned out at the tip, suggesting an effect of the gene on the nature of the material secreted. *Spineless* and *morula* slow down the growth of the bristle cells. *Singed*, *forked* and *Bristle* all affect the nature of the bristle secretion, there being some reason to suggest that the effects of *Bristle* and *singed* may be similar and different to that of *forked*.

1. INTRODUCTION

The numerous bristle mutants of *Drosophila melanogaster* have frequently been used in physiological experiments, but although embryological theories of their action have been freely offered (e.g. Goldschmidt 1931), data on the actual events which occur during the development of the bristles are almost lacking from the literature. Robertson (1936) drew attention to the presence of pairs of cells in the epidermis at about 27 hr. after puparium formation (at 25° C). One of these, that lying nearest the surface, will form the socket of the bristle and is known as the tormogen (Wigglesworth 1933), while the other, lying below, will form the bristle itself, and is known as the trichogen. Robertson also pointed out that the secretion of the bristle begins some time before any colour can be detected in it, and he indicated that the transparency of the uncoloured bristles had led some authors to overlook them and to suppose that they were not secreted until the much later time at which they first become visible in living pupae.

The purpose of this paper is to describe rather more fully the development of the bristles in normal and some mutant types of *Drosophila*. As is well known, there are two main types of bristles, the large macrochaetae and the small microchaetae. Most attention will be focused on the former, since many more mutations are known which affect them.

2. TECHNIQUE

The developing bristle cells were studied in serial sections and in whole mounts of the pupal epidermis. Whole mounts were prepared in the following manner. The pupae, which were timed at 25° C, were cut in two horizontally and fixed in Carnoy. After transference to 70 % alcohol the fat body was removed from the dorsal part of the thorax by means of a fine needle and the whole epidermis, after staining in Ehrlich's haematoxylin, was mounted while still adhering to the pupal cuticle. This method, providing a surface view of the developing cells, proved to be of special value when it was desired to establish the fact that certain pairs of bristle cells (e.g. in the mutant *scute*) never made their appearance.

Sections were cut at 8 μ , and were stained in Heidenhain's iron haematoxylin, or Delafield's haematoxylin counterstained with orange G.

3. THE DEVELOPMENT OF THE NORMAL MACROCHAETAE

The bristle-forming cells can at first be distinguished in the epidermis only by their size. This begins to become distinctly larger than that of the surrounding cells at about 15 hr. after puparium formation, that is, a few hours after true pupation. At this time, some of the cells appear to be single and others already in groups of two; thus the indications are that the final division is occurring at about this period. But this datum is by no means secure. It has proved impossible to identify the cells with certainty in whole mounts before about 19 hr., by which time they are invariably double, and in sections the absence of the second cell is not easy to confirm with absolute certainty. Probably experimental evidence, which is being accumulated and which will be published in a later communication, will prove a more convincing guide to this point than direct observation.

Both the trichogen and tormogen cells, from 16 hr. onwards, rather rapidly increase in size until by about 40 hr. they have some ten times the linear dimensions of the normal epidermal cells. In particular, the nuclei become very large, and contain deeply staining threads which show much the same structure as the salivary gland nuclei. There can be no doubt that these nuclei are polyploid, the increase in ploidy being due to endomitosis. At this time the nuclei also contain conspicuous nucleoli. In addition, the trichogen develops a massive, diffusely but fairly strongly staining cytoplasm.

The growth of the trichogen continues through the early period of the formation of the bristle. This begins at about 30 hr. or rather less, and proceeds rapidly; it appears to be complete by about 55 hr., but this is difficult to judge. The formation

of the socket appears to begin somewhat after the beginning of the secretion of the bristle. Wigglesworth (1933) claims that in the bug *Rhodnius* the bristle grows through a hole in the tormogen cell, which thus forms a complete socket for it from the very earliest stage. A similar state of affairs may have been described by Plotnikov (1904) in caterpillars of *Ocneria*, but his description is ambiguous. Wigglesworth claims the support of Hufnagel (1918), who, however, actually states that in the lepidopteran *Hyponomeuta* the bristle is at first enclosed only partially, resting in a slight concavity of the tormogen: and this is certainly so in the case of the scales of the lepidopteran wing (Stossberg 1938). It is also the condition in *Drosophila*, as far as can be made out. At least it may be said that if the tormogen is continuous all round the first protrusion of the bristle, the ring

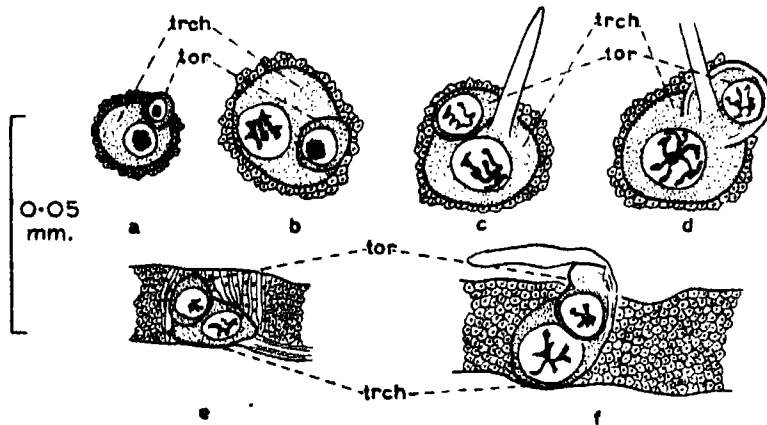


FIGURE 1. Development of wild-type bristles. *a*, surface view, whole preparation, about 22 hr.; *b*, ditto, about 28 hr.; *c*, ditto, about 31 hr.; *d*, ditto, about 40 hr.; *e*, section, about 22 hr.; *f*, section, about 31 hr., note secretion of tormogen (*tor*) does not completely enclose bristle secreted by trichogen (*trch*).

can only be completed by an extremely tenuous strand, and a careful search for such a strand has not been successful. The whole structure is very small in *Drosophila*, and it is difficult to rule out absolutely that the bristle emerges through a complete ring of tormogen material, but as yet there is no evidence for such a view. Moreover, when the tormogen begins to secrete the material which will eventually form the socket, the secretion is extremely asymmetrical as regards the bristle, being considerably thicker on one side, and thinning away in two arms which seem at first only partially to enclose the emerging bristle (figure 1). This secretion does not become noticeable until the bristle has already reached about half its final length, at about 36 hr. It has attained a considerable volume, and the socket has been completed all round the bristle, by about 48 hr.

The trichogen and tormogen secretions are both somewhat fibrous in microscopical appearance. The fibrous nature of the bristle substance is in the early stages indicated by a general weakly striated appearance throughout the whole

thickness of the bristle, which seems at this time to be a solid structure. In the adult the bristle is a hollow cone, and in cross-section one can see that the outer surface is raised into a number of shallow ridges which run from base to tip. The hollowing out of the interior occurs soon after the bristle reaches its full length, and is probably due to the hardening and contraction of the chitin of which the bristle is composed; the process is carried still further, and the conical wall of the hair becomes still thinner, when the bristles start to become coloured, at about 70 hr. The ridges can be seen as early as about 35 hr.

The trichogen and tormogen cells attain their maximum size at about 40 hr., when they are both actively secreting. At this time, not only are the nuclei very large, but there is a considerable quantity of cytoplasm, which stains more deeply in haematoxylin than the bristles do. This cytoplasm is, however, clearly destined to become the bristle, and as the latter grows, so the former disappears. As the secretion of the bristles and socket comes to an end, there is also a reduction in the size of the nuclei, and by the time the bristle begins to develop colour, the trichogen and tormogen cells have become merged into the epidermis and can only be distinguished with difficulty.

As Stern (1938) has pointed out, the bristles are associated with nerves. This connexion, which may be of great developmental importance, has not been followed in detail as yet, but is still under investigation. Under the larger bristles one can sometimes find a multicellular nerve strand, the cells of which have rather larger and paler nuclei than the cells of the epidermis.

4. THE DEVELOPMENT OF THE MICROCHAETAE AND CELL HAIRS

The development of the microchaetae does not appear to differ significantly, either in the processes involved or in their timing, from that of the macrochaetae. The microchaetal cells are throughout much smaller than those forming the macrochaetae, and are in consequence indistinguishable from the other epidermal cells at a time when the macrochaetal cells are already easily discernible. There is therefore even more doubt as to the exact time at which they divide, but the indications are that it is at about the same time as the macrochaetal cells.

The microchaetal groups of cells often appear to consist of three cells, arranged more or less on top of one another, with the middle one the largest. This largest cell is in fact the microtrichogen, and the smaller upper cell is the tormogen. The lowest cell is not a hair cell at all, but is a nerve cell, whose body and nucleus lies against the base of the trichogen, while a fibre leads away into the interior of the animal.

Still smaller than the microchaetae are the cell hairs, each of which is formed from a single cell. It is probable that every cell at the surface of the epidermis forms such a hair, just as do the cells of the wing membranes, but it is difficult to be certain of this. The hairs, which are very minute, can be made out at 50 hr., when the imaginal cuticle begins to be secreted, but have not been found earlier than this, although they may be present at earlier stages.

5. THE DEVELOPMENT OF *scute* BRISTLES

The numerous *scute* allelomorphs (*sc*, 1-0-0+) cause the absence of certain bristles. The alleles which we have studied are *sc* and *sc*², particularly the latter. Each *scute* allele affects a particular set of bristles, arranged in a certain pattern, but this regularity is only statistical; in any given individual it is impossible to be certain that a given bristle will be absent. In *sc*², however, there are rather regularly less than four scutellar bristles, and usually only two, sometimes only one.

The number of scutellar bristles has been investigated, both in whole mounts and in sections, at various stages of the pupa, and it has been found that in *scute* the number is less than four from the earliest stage at which the hair cells become recognizable. Since this is so, nothing can be said as to the manner in which they disappear, or the time at which the *scute* development first diverges from the normal course. Child (1935) found that the temperature-effective period for *scute*-1 lay at the end of the larval period. If, as would be expected, the developmental effects of this allele are essentially similar to those of alleles studied here, the action of the gene must occur in the few hours between the end of the larval period and the 20 hr. stage to which the bristles can be traced back in the pupa. Sturtevant (1932) has presented some evidence that *scute*, although autonomous in large regions of mosaics due to chromosome elimination in early developmental stages, is partially suppressed by wild-type tissue in small patches formed by late elimination. Diffusing 'bristle-forming substances' have freely been invoked to explain this and other facts, but one would like to know something of the anatomical relations with the peripheral nerves before adopting such hypotheses; investigations on these lines are being pursued.

Sturtevant & Schultz (1931) pointed out that in combinations of *scute* (and *achaete*, which is probably very similar) with *Hairless*, some bristles are more frequently removed than they are by either gene alone; in particular the heterozygote *sc*¹/*ac* is completely normal, while the combination *sc*¹/*ac*; *H* is considerably more denuded of bristles than is *H*. Unfortunately, they do not state whether the bristles are removed in a *scute* manner or in the quite different way characteristic of *Hairless*; it is clear that the two genes are reinforcing each other, but whether the *Hairless* is sensitizing the bristles to the action of *scute*, or vice versa, cannot be determined from the published facts. In males containing both *H* and *sc*¹, we have found that some bristles are removed together with the sockets, as in *scute*, while some are represented by double sockets, as in *Hairless*.

6. THE DEVELOPMENT OF *split* BRISTLES

The gene *split* (*spl*, 1-2.9) affects both the eyes and bristles; we are at present only concerned with the latter effect. The bristle phenotypes produced by *split* are actually several. Some bristles may be doubled, with a double socket, some may be single but thick, and also with a double socket, some may be single with a

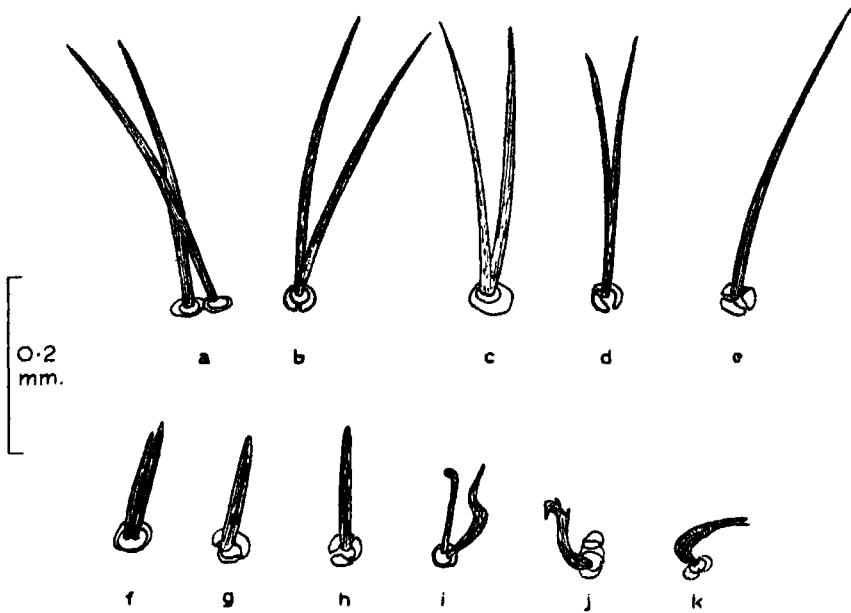


FIGURE 2. Bristles from adults of *split* (a-e), *split-Stubble* (f, g, h) and *split-forked* (i, j, k).

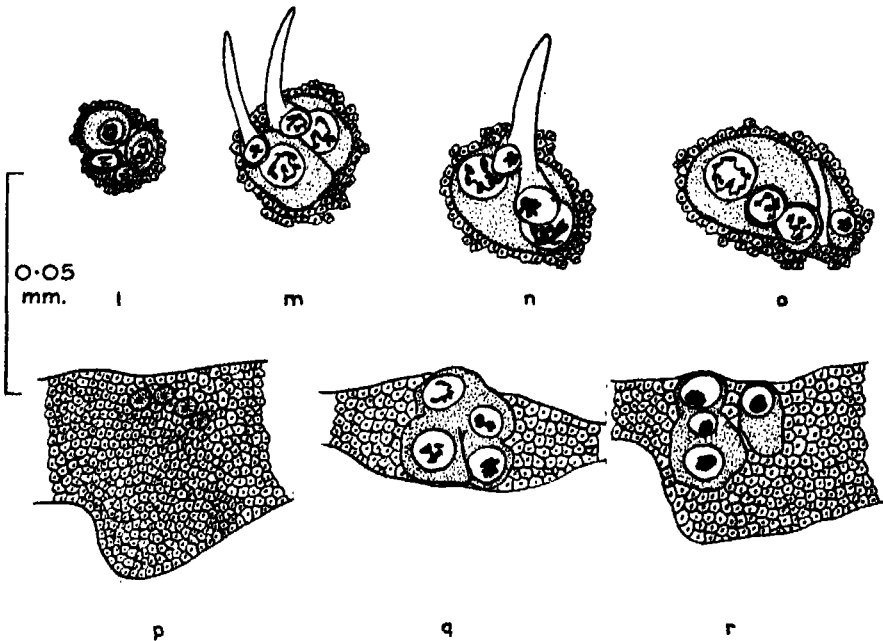


FIGURE 3. Development of *split* bristles. *l*, surface view, about 22 hr.; *m*, ditto, double bristle and socket, 31 hr.; *n*, ditto, single bristle and double socket, 28 hr.; *o*, ditto, single bristle and treble socket, 28 hr.; *p*, section, group of four bristle cells at an early stage, about 18 hr.; *q*, *r*, sections, about 28 hr.

tripartite socket, while some may be completely absent with no socket present either (figure 2 *a-e*). The different bristles on a single fly are usually differently affected, many remaining perfectly normal. The distribution of the effects among the different bristles has not been ascertained, but certainly varies in different stocks, and is thus under genetic control.

The histology of the developing *split* bristles is exactly what one might have guessed from a careful description of the adult. Instead of pairs of hair-forming cells, there may be groups of four (figure 3). It appears that an extra division of the cells occurs before their growth into trichogens and tormogens. In these groups of four, the most usual condition is for two cells, lying below, to become trichogens, while the other two lie above and become tormogens. In this case the two trichogens each secrete a bristle, but these bristles, if they lie in contact, may fuse, and appear as a single one. Similarly, the two sockets normally fuse to a greater or less extent.

This type of development is not invariable. In some cases three of the four cells lie on the surface and become tormogens, and there is then only a single bristle with a complex socket. In such sockets the three parts are usually rather symmetrically arranged; this provides additional evidence that the bristle does not grow through the substance of the tormogen cell, but becomes encircled by tormogen secretion. In other cases, and particularly (in the stock used) for the ocellar bristles, the groups of four cells may lie rather below the surface of the epithelium, in which case they remain rather small and do not secrete anything, either bristle or socket. The distribution of groups of four cells is therefore not so restricted as that of the double hairs in the adult, since many such groups produce only a single bristle or none at all. The actual distribution of the quadruple groups has not been determined.

In the combinations of *split* with some genes affecting bristle growth (*Stubble* (figure 2 *f, g*), *forked* (figure 2 *i-k*) and *singed*) the proportion of such irregularities as tripartite sockets is rather greater than in *split* itself. Presumably these genes have an effect on the regularity of the arrangement of the cells within the epithelium, and this reinforces the irregularity produced by *split*. Similarly, in the combination of *split* with *Hairless* (figure 4), the irregularity to which the effect of the latter gene is due is exaggerated by the disturbance produced by *split*, and this is true even in groups consisting of only two cells, which have therefore not undergone an extra division.

The time of the extra division caused by *split* has not been accurately determined, but it has already occurred by 21 hr. after pupation, which is about as early as the cells can be distinguished with certainty.

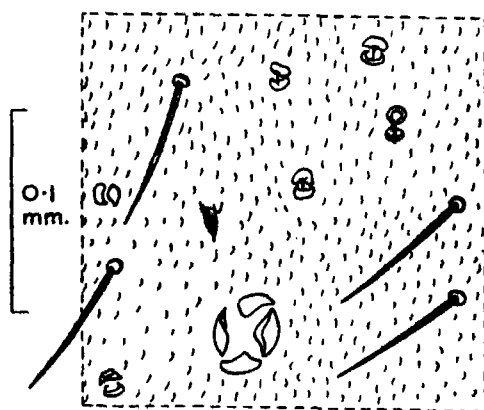


FIGURE 4. Part of thorax of adult of combination *split*; *H*.

7. THE DEVELOPMENT OF *singed* BRISTLES

The allelomorph *singed*³ (*sn*, 1-21.0) was investigated; it is one of the more extreme of the *singed* series.

The histological picture is from the beginning somewhat abnormal, although it is not always easy to state exactly the nature of the peculiarity. In many cases, however, the nuclear walls of both trichogen and tormogen, particularly of the latter, appear indefinite, so that the chromosomes seem to be loose in the cytoplasm of the cell. In other cases the tormogen appears flattened and distorted (figure 5).

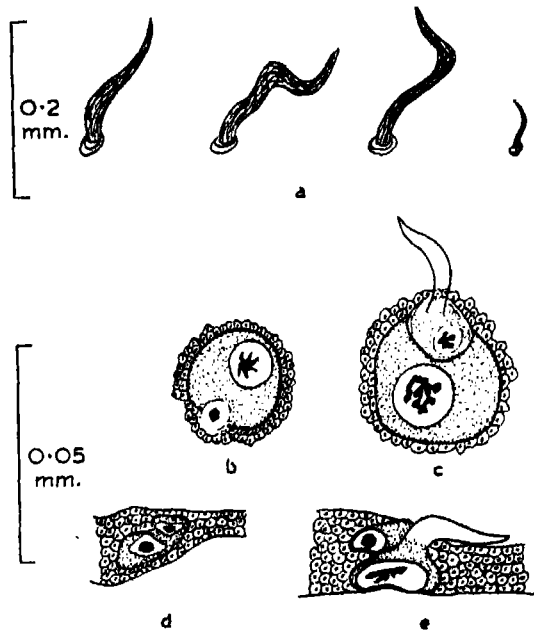


FIGURE 5. *a*, adult bristles of *singed*-3, one microchaeta to right; *b*, *c*, surface views of developing *sn*³ bristles showing thin-walled tormogen nuclei; *d*, section, about 22 hr.; *e*, ditto, about 31 hr.

Perhaps the indefiniteness of the nuclear boundary is due to a lack of the normal tendency for the cytoplasm near the nuclear membrane to stain deeply, a property which Caspersson (1941) suggests is generally related to the functioning of the nucleus as a seat of protein synthesis. In any case it seems to be true that the bristle secretion is abnormal, since the elongating bristles are curved from the very beginning. The greater abnormality of the nature of the bristle secretion in *singed* than in *forked* is also suggested by the fact that *singed* usually affects the small bristles (microchaetae and wing bristles) as well as the macrochaetae, whereas only the most extreme *forked* alleles do so.

In the combination of *forked* and *singed*³ the bristles are very strongly forked and bent (figure 6). The bases are often expanded, suggesting that in this case the

secretion is so abnormal that even the proximal part of the bristle buckles and swells out under the pressure of secretion.

In the combination of *singed* and *hairy* the effects of the two genes are additive; the supernumerary microchaetae on the scutellum produced by *hairy* are bent and twisted in the way characteristic of the ordinary microchaetae of *singed* flies.

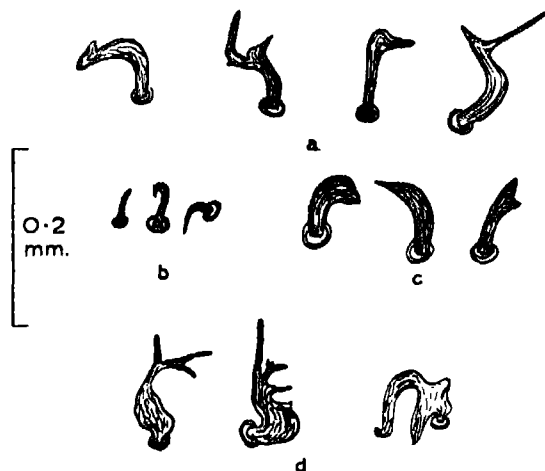


FIGURE 6. *a*, adult bristles of *forked*; *b*, adult bristles of *forked-spineless*; *c*, adult bristles of *forked-Stubble*; *d*, adult bristles of *singed-forked*.

8. THE DEVELOPMENT OF *forked* BRISTLES

Of the numerous *forked* alleles (*f*, 1-56.7), the two studied were *forked* and *forked*⁵. The histology of the developing bristles appears normal until some time after the secretion of the bristle has begun. At first the emerging bristles are perfectly straight and appear quite normal, but after growing to perhaps half their length they begin to exhibit curvatures and splittings near their tips.

Bristles of the *forked* type can be produced by operations made on pupae at a time when bristles secretion is proceeding; in the region of the injuries (made with a needle) the epithelium may adhere to the pupal chitin and produce mechanical distortions of the elongating bristles which mimic those characteristic of *forked*. Similar malformations have been found by Howland & Glancy (1937) in bristles developing in implanted imaginal disks, but in this case there is no way of telling at what stage of development the effect was produced. They also occur in mutants in which the proportions of the thorax are considerably modified, e.g. in a compound of *dachsous* and a so-called 'comb-gap complex', in which the whole fly is bloated, and parts of the thorax reduplicated, so that the space between the epithelium and the pupal chitin is reduced.

The degree of expression of *forked* is different in the various regions of the fly, being usually strongest in the scutellar bristles. It is also stronger in the macrochaetae than in the microchaetae; the latter are in fact usually normal except in

stocks showing a very high grade of *forked* effect. In combinations of *forked* and *hairy*, the microchaetae produced by the latter gene on the scutellum may be bent while the normal microchaetae on the thorax are unaffected.

It therefore seems that the process of bristle secretion is abnormal in *forked*. Somewhat similar abnormalities can be produced by mechanical hindrance to the process of secretion, but it is not clear that such hindrances occur in the development of the unoperated *forked* pupae. The fact that single, strongly forked bristles can be found in mosaics makes it probable that the *forked* phenotype can be produced in the absence of mechanical abnormality.

9. THE DEVELOPMENT OF *Bristle* BRISTLES

Bristle (*Bl*, 2-54·8) is a dominant gene which causes the bristles to be short and thin. On more careful examination it will also be seen that they have a typically beaded appearance, the outline being somewhat ragged, and the longitudinal

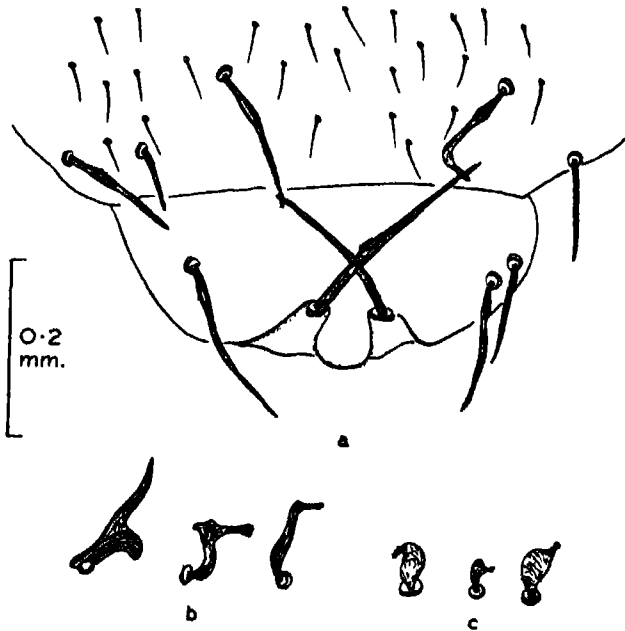


FIGURE 7. *a*, scutellar region of *Bristle* to show adhesion of macrochaetae and distortion of epithelium; *b*, adult bristles of *singed-Bristle*; *c*, adult bristles of *forked-Bristle*.

striation irregular. In the developmental stages, no abnormality of the trichogen-tormogen pair can be detected. There is, however, one significant fact about the appearance of the adult flies which may give a clue to what is happening. In the normal pupa, the posterior scutellars are directed obliquely forwards so that they cross one another and their tips lie adjacent to the posterior dorsocentral of the opposite side. In *Bristle* flies, very much more frequently than in other types,

bristles which cross in this manner may stick together. Their junction must be fairly firm, since when the epidermis expands in area towards the end of pupation, the bristles adhere strongly enough to cause considerable distortions (figure 7). It is difficult to avoid the conclusion that *Bristle* alters the nature of the secretion in some way which makes the bristles more sticky; the irregular and beaded appearance mentioned above suggests that the adherence of the bristles may be due to the interdigitation of minute projections from the bristle surface.

Bristle has been combined with *singed* and *forked*, both of which also affect the process of secretion. In the *sn³; Bl* flies the bristles are not markedly different from those of *sn³* alone, but *f; Bl* shows strong exaggeration, the bristles being quite short and inflated (see p. 100 for *Bl Sb*).

10. 'Extra-bristle complex 4le', AND THE DEVELOPMENT OF *hairy* BRISTLES

A gene, or genes, lying in the second chromosome and associated with dachshous-38*k*, causes the formation of extra dorso-centrals and sometimes of other extra bristles. The effects are rather similar to those of the 'extra-bristle complex' described by Sparrow & Reed (1940); as in their case, the exact genetic basis of the phenotype has not been determined.

Developmentally, the condition is, as it were, the opposite of *scute*; extra groups of bristle-forming cells are present in the epidermis from the earliest stages at which one could hope to detect them.

Except in the disposition of the extra bristle cells the development of *hairy* (*h*, 3-26.5) follows that of the 'extra-bristle complex' described above. The origin of the extra microchaetae on the scutellum (which normally bears only macrochaetae) is particularly easy to follow in whole preparations of the epidermis.

The combinations of *hairy* with *forked* and *singed* have already been mentioned (p. 95).

11. THE DEVELOPMENT OF *Dichaete* BRISTLES

The gene *Dichaete* (*D*, 3-40.4) causes a characteristic spreading of the wings, often brings about a partial failure of the two cephalic imaginal buds to fuse properly, and also produces several different effects on the bristles; we are here only concerned with the last of these effects. The commonest of the bristle effects is an absence of certain bristles, the anterior dorso-centrals and post-alars being most usually affected, the posterior dorso-centrals and scutellars more rarely; exact studies of the distribution of the effects have been made by Sturtevant (1918) and Plunkett (1926). When a bristle is absent, the socket is also missing. Another type of bristle effect was not uncommon in the *D/Ly* stock examined; certain bristles (particularly scutellars) may be cleft from the tip nearly to the base, forming a double prong, the two arms of which diverge at an acute angle (figure 8).

The developmental process leading to the latter condition is easily identified. In the epidermis of 48 hr. pupae it is sometimes found that the trichogen of a group has divided to form two smallish cells, while the tormogen remains single. Two sprouting bristles may be seen associated one with each of these trichogen nuclei. Presumably the single tormogen secretes a socket which encloses both these bristles, which are thus brought into such close contact that they fuse, forming the uncleft base of the bristle.

In cases in which a bristle is absent it is not so easy to distinguish the early stages in the abnormality. By 48 hr. a bristle which is going to be absent in the adult is already missing in the pupal epidermis. In rather earlier stages, at about 30 hr., it often appears at first sight that one bristle, for instance a right anterior dorso-central, is already missing; but a more careful inspection sometimes reveals the presence of a group of two or more cells which, although certainly not nearly so large as the normal trichogen-tormogen pairs at that time, yet seem slightly larger than the other epidermal cells. It may be that these groups represent the missing bristle-forming units, but the difference in size between them and the normal epidermal cells is so inconsiderable that it would be rash to lay too much stress on it. If such groups are accepted as abnormal bristle-forming cells, the phenomenon would be somewhat similar to that involved in the absence of bristles in *split*, where groups (of four cells, owing to an extra division) fail to reach the surface of the epidermis and to secrete. If, on the other hand, this identification of abnormal bristle cells is at fault, the elimination of the bristle by *Dichaete* must, like that caused by *scute*, occur at such an early stage that it cannot be traced histologically. It should be pointed out, however, there is an obvious similarity between the extra division of both trichogen and tormogen caused by *split* and that of the trichogen alone caused by *Dichaete*, and it is therefore tempting to suggest that the two genes also act in similar ways in eliminating bristles.

In the combination of *Dichaete* and *split* the adult flies are almost denuded of macrochaetae, the microchaetae being unaffected. The development of these flies has not been studied, but it seems most probable that the widespread disappearance of the bristles is to be taken as evidence that both genes suppress bristle formation by similar mechanisms, and reinforce one another in the compound. The failure of reinforcement in respect of the doubling of bristles could then be attributed to the fact that the mechanism causing a bristle to disappear completely acts at an earlier stage than that causing it to be doubled.

12. THE DEVELOPMENT OF *Stubble* BRISTLES

In flies carrying the dominant gene *Stubble* (*Sb*, 3-58.2) or homozygous for the allelomorphic recessive *stubbloid* (formerly *sbd*, now *Sb^r*), the bristles are shortened and to some extent thickened. They have, as near as can be judged, the normal volume, and the effect of the gene is mainly exerted on the shape of the bristle and not on its size (figure 9 a-c).

During the early part of pupal development, no abnormality can be seen in the bristle cells. At about 40 hr., that is to say, when the secretion of the bristle is proceeding rapidly, the relative positions of the trichogen and tormogen are slightly altered. The latter does not lie so directly above the former, and does not surround the base of the bristle so closely (figure 9 *f-i*). It appears that the abnormal thickness of the bristle is due to its being less closely invested by the tormogen during its secretion. It is noticeable that the greatest amount of relative displacement of the two cells is found on the dorsal side of the head which also corresponds to the position of the maximum expression of the *Stubble* character.

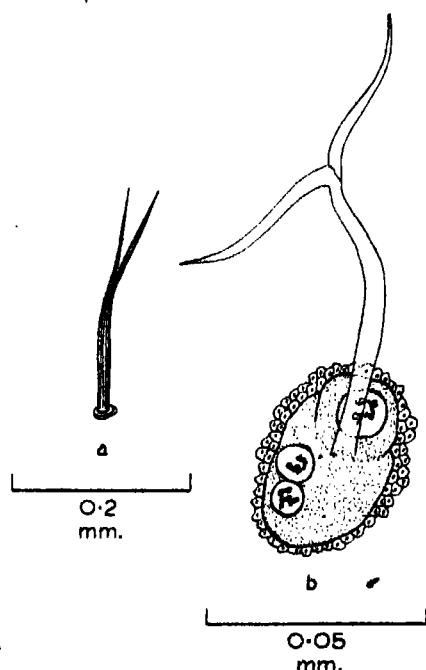


FIGURE 8

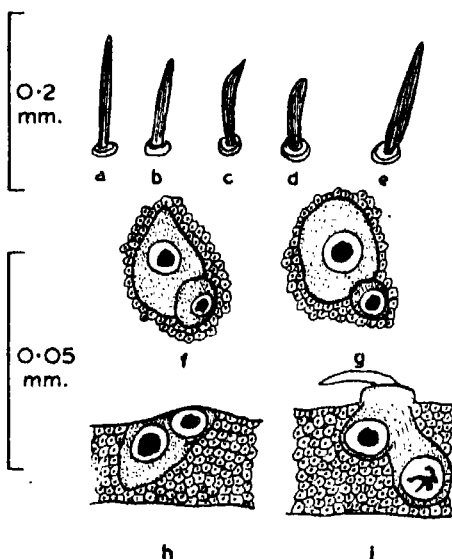


FIGURE 9

FIGURE 8. *a*, adult bristle of *Dichaete*, showing cleft form; *b*, surface view of developing cleft bristle, showing double trichogen nuclei on left.

FIGURE 9. *a*, *b*, *c*, adult bristles of *Stubble*; *d*, *e*, *Stubble*-like bristles produced near a wound made on the thorax of the developing pupa; *f*, surface view of developing *Sb* bristle in thorax region, tormogen lying slightly to one side of trichogen; *g*, ditto, from head region, the displacement of the tormogen being more marked; *h*, *i*, sections.

The abnormality in arrangement of the cells is not marked, and it is difficult to be quite certain of it. An alternative explanation for the effect of these genes might be that they affect the process of bristle secretion, acting in the opposite way to that which is suggested as the mechanism involved in *singed*. In the combination *singed Stubble*, however, there is no sign that the two effects tend to cancel one another out, as would be expected on this hypothesis; in fact the bristles are very

short and thick, and slightly bent. The same is true of *f*; *Sb* and *Bl Sb*. This militates against the possibility that *Sb* affects the process of bristle secretion, and falls into line with the view of its action advanced above.

The suggestion that the action of *Stubble* is primarily a mechanical one, caused by the abnormal relative positions of the trichogen and tormogen, is also supported by the fact that bristles of the *Stubble* type are sometimes produced by pupae in which the thorax has been injured by operation (figure 8 *d, e*). It is easy to imagine that such injuries may disturb the mutual orientation of the bristle cells, but it would not be simple to suggest a reason why such mechanical stimuli should alter the chemical nature of the bristle secretion.

The interaction of *Stubble* with *Hairless* and *split* is described on p. 102, and with *shaven* on p. 104.

13. THE DEVELOPMENT OF *spineless* AND *morula* BRISTLES

In *spineless* (*ss*, 3-58.5) both macro- and micro-chaetae are much reduced in size, although the effect is stronger on the former; the sockets are proportionately reduced. From what one can see in the pupae, the effect seems to be a simple one on the growth rate of the bristle-forming cells. These can first be distinguished at about the same time as in normal pupae, and are then of the normal size. They fail, however, to increase in volume at the normal rate, and, since they begin to secrete at the usual time, they are by then considerably smaller than the bristle cells of wild type. This slowness of growth characterizes not only the cytoplasm but also the nuclei of the cells. It is not uniform over the whole fly, since the scutellars, for example, are proportionately more reduced than the dorsocentrals.

In *morula* (*mr*, 2-106.7) the bristles are reduced in size in a way very similar to that characteristic of *ss*, and the developmental appearances are very similar.

In flies showing *spineless* combined with *split* the effects of the genes are simply added together, with no marked interaction. *Spineless* has not been combined with *morula*. With *forked* and *singed*, *spineless* causes a marked decrease in the abnormality of the bristles which are usually merely bent, and very rarely exhibit branching or forking. The *spineless-forked* macrochaetae (figure 6*b*) in fact resemble normal forked microchaetae, and it thus seems probable that the effect of *forked* (and to a lesser extent *singed*) is dependent on the absolute size of the bristles.

14. THE DEVELOPMENT OF *Hairless* BRISTLES

Hairless (*H*, 3-69.5) is a dominant gene which causes the absence of certain bristles. The sockets of the missing bristles are present and can easily be seen with a binocular microscope at the magnifications normally employed for classifying *Drosophila* families. The presence of the sockets originally led to the suggestion (Bridges & Morgan 1923) that in *Hairless* the bristles are at first formed but later

become detached. This is not the case. The effect of *Hairless* is essentially to cause both the hair cells to develop as tormogens.

In the stocks used, the bristles whose absence could be most fully relied on were the posterior ocellars, and the observations were primarily based on observations of their development; but once the *Hairless* effect had been identified, other bristles could be seen to be affected in the same way. In places where *Hairless* is effective, the two hair cells lie side by side instead of one on top of the other, as they do in normal hairs (figure 10 *e-i*). From the two cells thus formed, two masses of rather shapeless secretion are produced, and form two elevations capping the cells. It is these two elevations which later appear as a socket; examination of the adult fly with a compound microscope, giving a higher magnification than the usual binocular, will immediately show the atypical structure of the sockets (figure 10 *a-d*).

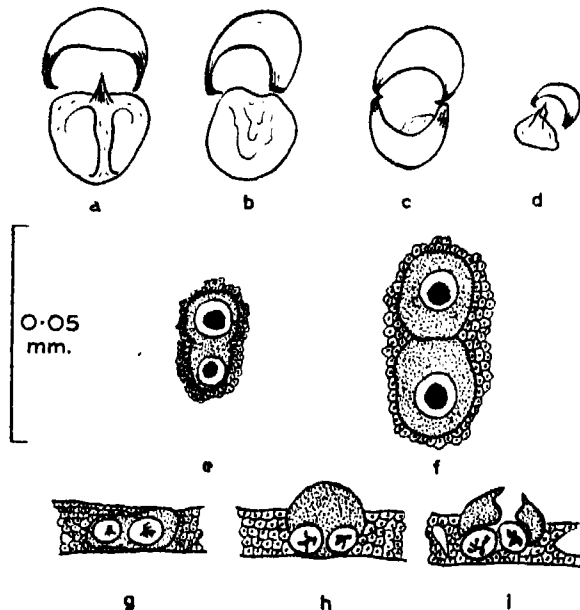


FIGURE 10. *a*, 'missing bristle' of *Hairless* adult, note minute spike of bristle; *b*, *c*, *Hairless*, both cells produce only socket; *d*, microchaeta of *Hairless* adult; *e*, surface view of developing *H* bristle, about 22 hr.; *f*, ditto, about 31 hr.; *g*, section, about 22 hr.; *h*, *i*, sections, about 40 hr.

The side-by-side arrangement of the two hair cells has been traced back as far as the cells are distinguishable, and probably originates soon after the last division of the cells. Whether it is caused by the division occurring across a vertical instead of a horizontal plane, or whether the cells migrate to their abnormal positions after a normal division, cannot be stated.

The relative position of the two cells is not quite regular, and in some cases the trichogen is only partly converted into a tormogen; that is to say, it eventually secretes a small amount of bristle, which protrudes as a tiny sharp projection between the two masses of socket material.

In combinations of *Hairless* with *split* (figure 4) the fly is very extensively denuded of hairs, both macro- and micro-chaetae being affected. Quadruple sockets are formed rather frequently, but the effect cannot be explained merely as a result of *Hairless* causing an orientation of the four cells less favourable for bristle formation, since there are many more absent bristles in *H; spl* than there are groups of four cells in *spl*. Most of the missing hairs are in fact replaced not by quadruple sockets but by double sockets. It seems that the reinforcement is not so much in relation to the occurrence of an extra division, but rather in respect of the orientation of cells which have never divided more than once. The tendency of *split* to cause abnormal orientation usually only comes to light in those cells which divide twice, but there must be a subthreshold effect on the other groups of hair cells which is reinforced by the similar tendency of *Hairless*.

The exaggeration of the hair elimination characteristic of *spl; H* flies is not found in combinations of *Hairless* with *Stubble*, although we have seen (p. 100) the effects of the latter gene can be exactly imitated by mechanical means, and there is every reason to believe that the gene acts by altering the arrangement of the two bristle cells. In *H Sb* flies a few more bristles may be missing than in *H* ones, but the effect, if any, is so slight that it could only be made certain of by careful statistical analysis of isogenic stocks. The bristles which remain present are typically *Stubble* in character, their shortness and thickness not being exaggerated in any way by the presence of *Hairless*. One must presumably conclude that the disorientations produced by *Hairless* and by *Stubble* are brought about by quite different mechanisms, so that the presence of *Stubble* does not make the complete side-by-side arrangement of *Hairless* more probable nor the presence of *Hairless* increase the degree of disorientation in the majority of the bristles.

The suggestion that *Hairless* and *Stubble*, although causing somewhat similar abnormalities in the arrangement of the cells, do so by different mechanisms, is supported by the fact that *Stubble*, unlike *Hairless*, is unaffected by the simultaneous presence of *split*. In *spl; Sb* flies there are rather more abnormalities of the *split* type (i.e. double bristles, tripartite sockets, etc.) than in pure *split* ones; but the bristles remain characteristically *Stubble* in appearance, even when two emerge from a double socket. Bristle-forming groups containing *Sb* must in fact have a tendency to arrange themselves in a certain equilibrium position (which may be slightly different in different parts of the fly) which is such that a bristle of the characteristic shape emerges.

15. THE DEVELOPMENT OF *shaven-naked* BRISTLES

The allelomorphs of *shaven* (IVth chromosome) reduce the length of the bristles. The allele studied was *shaven-naked* (*svⁿ*), which has a strong effect, which is somewhat dependent on the temperature at which the flies are reared, being slighter at 25° C than at room temperature. On careful examination it will be seen that in the adult fly the sockets are very large, and rather irregularly shaped. In some

places no bristle is formed at all, but more usually a more or less reduced bristle emerges rather to one side of the socket (figure 11*b*). The shorter bristles are usually very peculiar in shape, with the tip spread out into a number of very fine projections (figure 11*a*); longer bristles may be more regular, but in some cases they are shaped like a T with a very short stem; almost immediately above the socket, they divide into two portions which extend in opposite directions (figure 11*c*).

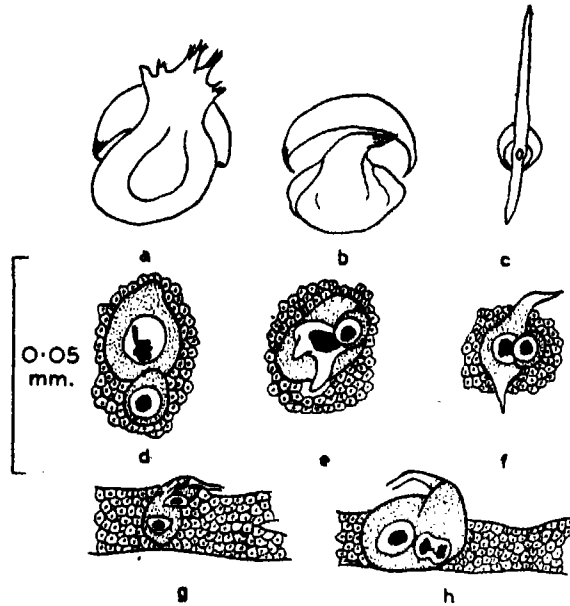


FIGURE 11. *a, b*, extreme types of adult bristles in *shaven-naked*; *c*, T-shaped bristle in *sv**; *d, e, f*, surface views, *d* showing side-by-side arrangement of tormogen and trichogen cells, *e*, showing lobular nuclei, and *f*, the secretion of a T-shaped bristle; *g, h*, sections.

From this description it is clear that the *shaven* phenotype resembles *Hairless*, in that the reduction of the bristle is not accompanied by a proportionate reduction of the socket. On the other hand, in *Hairless* there is a more or less complete substitution of extra socket secretion for the bristle in certain places, while other groups of hair-forming cells develop normally, whereas in *shaven* the effect, although somewhat variable in the different bristles, is by no means an all-or-none affair, and nearly all bristles on a fly are affected to some degree. Moreover, the extensive fanning out of the tips characteristic of *shaven* has not been found in the tiny bristles which sometimes occur in *Hairless*.

The development of *shaven* also clearly shows a resemblance to that of *Hairless*, in which, as was pointed out above, the main effect is an alteration in the relative positions of the trichogen and tormogen cells, so as to bring them side by side instead of one above the other. A similar abnormal position is taken up by the hair cells in *shaven*, but, although some shift in position is found in every group of cells, the degree of shift is not as constant, or usually as considerable, as in *Hairless*.

(figure 10 *d-h*). This is, of course, just what would be expected from the adult phenotype. It has not been possible to determine with certainty whether in *shaven* the trichogen first secretes a length of hair, and then passes over to the secretion of socket, or vice versa, but the evidence favours the first alternative. The tendency towards a side-by-side arrangement of tormogen and trichogen is, however, not the only abnormality which can be seen in the developing hair cells of *shaven*; in both cells the nuclei are peculiarly lobular, so much so that in some cases one might think that more than two nuclei were present. The significance of this irregularity in shape of the nuclei is not clear, but it is not unlikely that it is an indication of some peculiarity in the secretion process. That the character of the secretion is affected by the gene is strongly suggested by the spiky tips of the hairs which have been mentioned above.

The observational evidence thus suggests that in some respects at least, *shaven* is nearly related to *Hairless*. Flies exhibiting the combined effects of both genes show that the processes concerned are similar in nature, since they very strongly reinforce each other (figure 12). The macrochaetae are almost completely absent, their place being taken by double sockets of the *Hairless* type, with scarcely a sign of a bristle, however minute. Even many of the microchaetae are affected in the same way. The bristle reduction is even more extreme than in *Hairless-split*; and of course the quadruple sockets are quite absent in *Hairless-shaven*.

The combination of *shaven* with *Stubble* is rather similar to *Hairless-shaven*, particularly as regards the macrochaetae. The microchaetae are not quite so strongly affected, since instead of being transformed into mere double sockets, they are for the most part in the form of T-shaped hairs, with a short stem, and two oppositely directed arms. The mechanism by which these T-shaped hairs and bristles are formed is not at all clear.

In both *H; svⁿ* and *Sb; svⁿ* the sex combs of the males may be affected. The size of the teeth (which are really thick bristles) is irregularly reduced, some of them being almost completely absent. A casual inspection, however, gives an exaggerated impression of the number of teeth missing, since many of the shortened ones fail to darken properly, and remain almost transparent. This is probably an indication that they are not hollow and are therefore less permeable to the tanning and darkening substance; and, like the thickening of *Stubble* bristles, the effect can probably be attributed to atypical arrangement of the two generating cells.

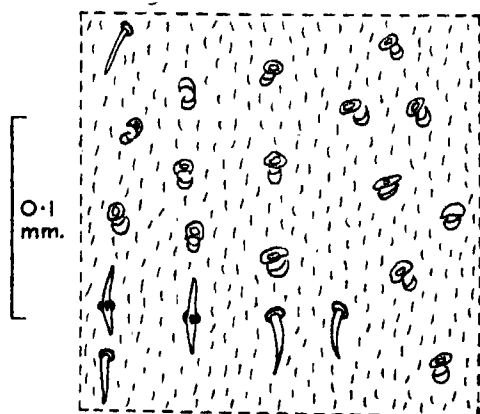


FIGURE 12. Part of thorax of adult of combination *H; svⁿ*. Only microchaetae are shown.

16. DISCUSSION

Our concern in this paper has been to discover what Ephrussi (1939) has called the 'point of attack' of the various genes investigated. It is important to understand just what aspects of developmental physiology can be elucidated by such studies. The final pair of bristle cells is formed as the end-product of a series of cell divisions which reach back to the fertilized nucleus of the zygote. If we compare two bristle pairs of different genotype, then those genotypic differences have characterized the nuclei of all the cell ancestors of the two pairs; the genetic difference between a *split* hair group and a wild-type hair group, for example, originated at the time the zygotes were first formed, and has persisted from that time until development is finished. But the two groups of hair cells, although different genetically, and therefore chemically, may undergo similar changes for a considerable part of their early development. During such stages of similarity, the gene, e.g. *split*, is therefore not demonstrably active. The 'point of attack' is the first time at which a cell of a given genotype can be shown to differ from the wild type in anything except the mere presence of an abnormal genetic constitution. The first provable divergence may be in the chemical constitution of the non-genetic constituents (e.g. cytoplasm, nuclear sap, etc.) or it may be morphological, as is usually the case in the present investigation.

Clearly it will always be difficult to make certain that one has discovered the actual original point of attack of a gene. The experimental datum is always in the form of a determination of the earliest time at which an effect of the gene can be discerned by the particular methods employed; other methods of investigation might be able to reveal earlier stages of divergence from the normal course of development. But the study of the points of attack of genes is in a very elementary stage, and new data are not without interest even if subject to later amplification and revision.

In the present investigation, the points of attack of twelve genes affecting *Drosophila* bristles have been determined, at least in a preliminary way. They are summarized in the diagram (figure 13). Such a diagram immediately suggests several points which require discussion.

In the first place, some of the genes are indicated more than once. A single gene has, in fact, several points of attack. Thus *split* is shown as attacking the division of the hair cells and their arrival at the surface of the epithelium. Such a statement may be merely a confession of failure to determine a single fundamental point of attack. It is possible that it is one and the same reaction of the *split* gene which frequently brings about an extra division of the cells and their occasional failure to reach the surface of the epithelium; and it may be that this same reaction is also concerned with the effects of *split* on the differentiation of the ommatidia of the eye. But if that is the case, it must be stated that this master reaction cannot yet be identified. And it is, with our present knowledge of gene actions, still perfectly possible to suppose that the gene, in bringing about these different effects, is

performing different reactions, by interacting with different substrates. It may be that the different processes affected by a gene with many different effects (a 'pleiotropic' gene) have nothing in common except the intervention of that gene. We must therefore be willing to accept the possibility of their being several points of attack of the same gene on different developmental pathways, although we should, of course, persist in attempts to discover features in common between the different gene reactions. But for many genes we must be content with a description of its pattern of attack points.

Stages of development	Genes
Determination of bristle cell	<i>sc, eb^{41a}, h</i>
↓	
Division	<i>spl, D</i>
↓	
Arrival at surface	<i>spl, D</i>
↓	
Arrangement of trichogen and tormogen	<i>spl, H, Sb, sv</i>
↓	
Growth	<i>ss, mr</i>
↓	
Secretion	<i>sn, f, Bl, sv</i>
↓	
Darkening and hardening	<i>sn, f (at kinks)</i> <i>H; sv, Sb; sv (sex combs)</i> <i>y, e, stw, b</i> and other 'body-colour' genes

FIGURE 13. Points of attack of various genes on the development of the bristles.

The difficulty of identifying one single fundamental gene effect is particularly great for those genes which affect other organs as well as the bristles. Thus *Hairless* frequently causes a shortening of some of the longitudinal veins of the wing, and *Dichaete* causes the wings to be held out at an angle from the sides of the body; no connexion between these effects and the hair abnormalities can yet be suggested.

Spineless is an allelomorph of the well-known *aristopedia* series, other members of which have the remarkable effect of transforming the antennae into leg-like organs. The connexion between such an effect and the slowing of hair growth caused by *spineless* itself is not obvious at first sight; but Waddington (1942) has shown that changes of developmental fate somewhat similar in kind to that caused in the antenna by *aristopedia* may be brought about by abnormal growth, and it therefore seems conceivable that the two effects of the *spineless* locus may eventually turn out to be less disconnected than they seem.

Very similar to *spineless* in its effect on the bristles is *morula*, another gene which has effects on quite other organs. But in this case it is the eye which is affected. The abnormalities produced seem (Waddington & Pilkington, studies in progress) to be almost identical with those caused by another gene which affects both eyes and bristles, namely, *split*. The eye effects of this gene have been described in detail by Pilkington (1941). They can, like the hair effects, be regarded

as consequences of abnormal cell arrangements and cell divisions; but in the eye there are too few divisions, whereas in the hairs there are too many. Much further work will be necessary before we reach a full understanding of the pleiotropic activity of these two genes, and of the reasons why they produce such different effects on the hairs though such similar ones on the eyes.

A considerable number of gene combinations have been studied in this investigation and are mentioned or described in the earlier sections. In discussing the interaction of genes, very little attention has been paid in the past to their points of attack, probably because information on this point was not available. Geneticists have distinguished pairs of genes whose effects are merely summated in the combination, those which reinforce or inhibit one another to a greater degree than mere summation, and those in which one gene masks the effect of (or 'is epistatic to') the other. The mechanisms by which such interactions can be brought about will depend to a considerable extent on the similarity or difference of the points of attack of the genes, and when the points of attack have been determined, the categories mentioned above will be found to be portmanteau groups, containing very mixed assemblages of gene interactions. It appears more profitable at the present stage to discuss the interactions of the bristle genes from first principles, without attempting to classify them into rigidly defined categories.

In many cases a fly homozygous for two genes shows the summated effects of both genes with no sign that the presence of one gene has altered the expression of the other. This may be true either of genes with different points of attack (e.g. *sn*; *h*, *f*; *h*) or with the same point of attack (e.g. *spl*; *Sb*). Such uncomplicated summation is only to be expected in the former case, since the genes are presumably doing quite different things, and there is no reason why they should affect each other. But it is somewhat more unexpected for two genes which share a point of attack, and when it occurs in such cases, it presumably indicates that the two genes, although attacking the same phase of development, act on different processes proceeding at that time. For instance, consider the *split-Stubble* combination. There is no reason why the presence of *Stubble* should alter the effect of *split* on the division of the bristle cells, and in fact it does not do so. But *split* also tends to cause a disorientation of the cells, while *Stubble* causes the tormogen to lie less directly over the trichogen. Now in both these genes, equilibrium positions of the two cells are involved; in *split* there must always be some tormogen which bears a normal relation to the trichogen, so that the bristle secreted is of normal shape, and in *Stubble* the relative positions of the two cells are modified, not at random, but in a regular way so as to produce bristles of the characteristic length and diameter. In the combination, the bristles are of the usual *Stubble* type, that is, the trichogens are accompanied by tormogens lying in the relative position normally produced by the *Stubble* gene. The slight increase in the proportion of treble sockets with a single hair as against double sockets and hairs, mentioned on p. 102, was no greater in the *split-Stubble* combination than in other cases in which *split* had been out-crossed, and probably did not represent an effect of the *Stubble* gene. We can

therefore describe the *split-Stubble* combination by saying that *split* produces its usual effect of causing the formation of irregularly arranged groups of four cells, while *Stubble* also has its normal effect of modifying the relative positions of tormogens to their related trichogens.

Similar additive effects are found in the *Hairless-Stubble* combination. Threshold effects are very clearly involved in the *Hairless* phenotype; the trichogens either lie in the normal position relative to the tormogens, giving rise to normal bristles, or they lie beside the tormogens, when they form no bristle at all but a second socket. In *Hairless-Stubble* the trichogens may lie beside the tormogens, giving a double socket, or they may take up the *Stubble* position, and form a thick bristle. But in those bristles in which *Hairless* does not produce its effect, there seems to be no subthreshold effect which can modify the *Stubble* equilibrium; the bristles are not noticeably shorter or thicker than in normal *Stubble*.

When we examine the *split-Hairless* combination we no longer find a mere addition of the two effects, but rather a marked enhancement of the *Hairless* phenomenon and some increase in the frequency of the extra divisions caused by *split*. The latter effect is probably no greater in this case than in other outcrosses of *split*, but the former requires some explanation. This may not be too difficult to provide when one realizes that the effects of *split* and *Hairless* on cell orientation are essentially of the same kind. *Hairless* causes the trichogen to lie either normally with respect to the tormogen or beside it; *split* causes a trichogen either to have the normal relative position or to become a tormogen, but the latter possibility is only realized when the trichogen is one of a group of four cells produced by an extra division. But any subthreshold effect there may be in the two-cell groups towards such an orientation will not, when in combination with *Hairless*, have to modify an equilibrium position, but will merely increase the frequency with which the abnormal one of the two possible equilibria is attained.

The additive rule also breaks down in combinations between either *Stubble* or *Hairless*, both of which are characterized by threshold phenomena, and *shaven-naked*, in which there is clearly no stable equilibrium relation between the trichogen and tormogen. Here we could say, if we liked, that the *shaven* absence of equilibrium is epistatic; in both combinations most bristles are reduced to minute stumps or are converted completely into sockets. The formation of numerous T-shaped microchaetae by *Stubble-shaven* is peculiar; it is not yet perfectly clear whether this type of bristle is a consequence of a particular spatial disarrangement of the two cells, or whether it is a result of some peculiarity of nature of the bristle secretion.

Several different varieties of interaction are also found in combinations of genes affecting bristle secretion. These mostly behave, as would be expected, in a simply additive way in combinations with genes with a different point of attack, e.g. *Bristle-Stubble*, *singed-Hairless*, etc. One could, perhaps, claim that *Hairless* is epistatic to genes such as *singed* in those bristles which the former reduces to a double socket; but except as an example of the complexities arising out of a

mechanically applied classification, such a remark would not add anything to our understanding.

When combined with each other, the bristle-secretion genes behave in a way which could hardly be predicted. *Shaven* has not been combined with the other three. *Forked*, when combined with *singed* or *Bristle*, shows exaggeration, while the *singed-Bristle* compound looks nearly indistinguishable from *singed*. No thresholds are apparently involved and two hypotheses are possible in explanation of the phenomena. First, one might suppose that all three genes act through different chemical mechanisms; in such an event one would expect addition or exaggeration in all combinations, and the lesser effect of *Bristle* in *sn; Bl* than in *f; Bl* would require special explanation. Or secondly, one could suppose that *Bristle* and *singed* act in the same way and *forked* differently. We should then have exaggeration in the combinations of the two different actions (*f sn* and *f; Bl*); and the combination of the two similar ones (*sn; Bl*) could probably be regarded as exhibiting addition, since the effect of *Bristle*, if of the same kind as that of *singed*, is of such a slight grade that it would be scarcely noticeable in a simple addition.

There are one or two particular points to which attention should be drawn. Schultz (1934) has shown that *shaven* behaves like a hypomorph in Muller's sense; the length of bristle produced depends on the number of *shaven* genes present. With our present knowledge of the mechanism of the *shaven* reaction, this result is perhaps rather less easy to picture than it would be if one could suppose that *shaven* is directly concerned with the synthesis of hair substance. One has, now, to suppose that the gene produces some substance which tends to make the trichogen lie normally below the tormogen and thus to continue secreting a bristle rather than a socket. Schultz also stated that tetra-IV diploids containing four doses of *shaven* exhibit extra bristles, and he took this as evidence that sufficiently large doses of a hypomorphic gene may produce more of its effect than does the wild-type allele. Such a phenomenon would not be easy to understand, since it is clear that two doses of the wild allele are usually no more effective than one; moreover, in other cases, such as the *bobbed* alleles studied by Stern (1929), it was found that the wild-type level of production was never exceeded, however great the dosage of hypomorphs in the genotype. It appears from the present account that these difficulties do not in fact arise, since it is clear that the formation of extra bristles cannot be regarded as in any way an exaggeration of the normal *shaven* process. It is difficult to know what form such an exaggeration might take—perhaps the transformation of socket into bristle—but the formation of extra bristles must have been due to some quite different process, perhaps a result of the generally unstable developmental processes of a tetrasomic animal.

The question of what actually happened in Sturtevant and Schultz's *Hairless-scute* flies has been raised above, and cannot yet be answered.

REFERENCES

- Bridges, C. B. & Morgan, T. H. 1923 *Publ. Carneg. Instn.* no. 327.
Caspersson, T. 1941 *Naturwissenschaften*, **29**, 33.
Child, G. P. 1935 *Genetics*, **20**, 127.
Ephrussi, B. 1939 *Génétique Physiologique*, Paris.
Dobshansky, T. 1927 *Z. indukt. Abstamm.- u. Vererb. Lehre*, **43**, 330.
Goldschmidt, R. 1931 *Biol. Zbl.* **51**, 507.
Howland, R. B. & Glancy, E. A. 1937 *Amer. Nat.* **71**, 429.
Hufnagel, A. 1918 *Arch. Zool. exp. gén.* **57**, 47.
Pilkington, R. W. 1941 *Proc. Zool. Soc. Lond.*
Plotnikov, W. 1904 *Z. wiss. Zool.* **76**, 333.
Plunkett, C. R. 1926 *J. Exp. Zool.* **46**, 181.
Robertson, C. W. 1936 *J. Morph.* **59**, 351.
Schultz, J. In Morgan, Bridges & Schultz (1934). *Yearb. Carneg. Instn.*, no. 33.
Sparrow, A. H. & Reed, S. C. 1940 *J. Genet.* **39**, 411.
Stern, C. 1929 *Biol. Zbl.* **49**, 261.
Stern, C. 1938 *Genetics*, **23**, 172.
Stossberg, M. 1938 *Z. Morph. Ökol. Tiere*, **34**, 173.
Sturtevant, A. H. 1918 *Publ. Carneg. Instn.* no. 264.
Sturtevant, A. H. 1932 *Proc. 6th Internat. Congr. Genet.* **1**, 304.
Sturtevant, A. H. & Schultz, J. 1931 *Proc. Nat. Acad. Sci., Wash.*, **17**, 265.
Waddington, C. H. 1942 *Nature, Lond.*, **149**, 264.
Wigglesworth, V. B. 1933 *Quart. J. Micr. Sci.* **16**, 269.

CROONIAN LECTURE

Chromatic behaviour

BY LANCELOT HOGBEN, F.R.S.

(Delivered 18 June 1942—Received 18 June 1942)

Visible colour changes of animals are of two types, those which result from build-up or breakdown of pigments, and those which result from their redistribution. Assumption of winter dress by some mammals and of breeding dress by some fishes are familiar examples of the former. The proverbial behaviour of the chameleon illustrates the latter. Mere redistribution of pigment may itself come about in several ways. It may result from the responses of contractile tissues *sensu stricto* or from those of pigmentary effector organs *sui generis*. Visible colour changes due to the response of contractile tissues include: (a) the familiar phenomena of blushing, blanching and cyanosis due to surface variation of the quantity of blood pigment, i.e. to vasodilation, vasoconstriction and capillary stasis; (b) the more specialized chromatic behaviour of cephalopods. Cephalopod colour change is brought about by relaxation and contraction of radially disposed plain muscle fibres attached to pigment-containing vesicles. What follows refers only to chromatic behaviour accomplished by effector organs other than muscle fibres. At present we know of two main types:

(i) The pigmentary effector system of vertebrates, of Crustacea and of Annelida* (Wells 1932) consists of cells in which pigment granules or droplets migrate to the periphery of diffusely branching processes, thereby offering more surface for absorption of radiation.

(ii) The pigmentary effector organs of Culicidae, at least those of the air sacs of *Corethra* (Martini & Achundow 1929), are migratory cells capable of wandering to situations where they offer more or less obstruction to surface transmission or reflexion of incident light.

There is relatively little available information about the structural basis of chromatic behaviour in other insects, e.g. phasmids. Apparently it is brought about by clumping or dispersion of pigment granules or droplets in the cells or syncytium (hypodermis) immediately beneath the cuticle. The immediate physical basis of chromatic behaviour is essentially as in vertebrates, Crustacea and annelids; but there are no distinctive structural elements answering to the pigmentary effector system of these groups. To date, the volume of published work on insects is meagre. One reason for this is lack of a satisfactory method for recording and assessing colour change. Apart from rough and ready visual impressions of skin appearance or merely qualitative specification of the condition

* See also Lebour, M. V., *Nature, Lond.*, 150, 209 (1942).

of the pigmentary effector organs, research on chromatic behaviour of vertebrates or of Crustacea can take advantage of three methods for recording it: (a) photoelectric recording (Hill, Parkinson & Solandt 1935) of the whole organism; (b) micrometer measurements (Wykes 1936) of the mean diameter of pigment-spread in single chromatophores; (c) a scale of arbitrary numerical values (Hogben & Gordon 1930; Hogben & Slome 1931) assigned to easily recognizable grades of pigment dispersion in *living* pigment cells.

A sufficient objection to (a) is that the pigmentary effector system of vertebrates and of Crustacea usually includes several different types of cells, visibly distinguished by size and situation (e.g. dermal and epidermal melanophores of Amphibia and fishes) or by the pigments they contain—brownish black (melanophores), red (erythrophores), yellow (xanthophores), iridescent white (leucophores). The melanophores, erythrophores, xanthophores, etc., of most Crustacea, like those of vertebrates, are scattered independently at random in the appropriate region of the skin. Decapods of the order Natantia have compound chromatophores, i.e. each erythrophore or melanophore is closely united to a xanthophore and sometimes to a leucophore. Since chromatophores of different types do not respond in the same way or at the same speed to different stimuli, similar physical responses recorded by the photoelectric method may correspond (Neill 1940) to totally different physiological phenomena.

Another objection to the photoelectric method is the fact that chromatic behaviour and structural colour change may go hand in hand (Dawes 1941). External circumstances may lead to the formation or elimination of pigment cells in some Amphibia, e.g. *Xenopus** and in some fishes. Metabolic changes involving the total quantity of pigment and mechanical changes affecting its distribution seem to be inextricably involved in the colour responses of phasmids (Giersberg 1930); and there is reason to believe that appearance of blue colouring matter in their cell processes during aggregation of the erythrophores of some shrimps, e.g. *Hippolyte* (Keeble & Gamble 1904) and *Palaemonetes* (Brown 1935), accompanies breakdown of the red pigment *pari passu*.

Thus a satisfactory method of assessing chromatic behaviour *sensu stricto* must direct attention to the condition of individual pigmentary effector organs. A disadvantage of the second method (b) is that it is time-consuming. Hence it is not suitable for study of rapid response without recourse to fixation. Fixation has a double disadvantage. It is uneconomical and it leads to distortion. Fully dispersed melanophores undergo slight aggregation at death, and fully aggregated melanophores undergo slight dispersion. Consequently, fixed material limits our range of observation. Moreover, the micrometer procedure is satisfactory only for chromatophores of highly regular configuration such as those of Teleostei. The third method is not open to these objections, and yields remarkably consistent results in the hands of different observers. As prescribed (Hogben & Gordon 1930) for melanophores, the scale has five grades ranging from 1.0 (complete aggregation)

* Personal observation.

to 5.0 (maximal *dispersion*). The symbol μ stands for the mean value (melanophore index) of at least twelve individuals subjected to the same conditions.

An immediate obstacle to study of the mechanics of the pigmentary effector organ is lack of unanimity concerning the sequence of events. Application of micro-dissection technique to melanophores of *Fundulus* (Matthews 1931) supports the consensus of testimony from recent workers, i.e. that pigment diffuses within fixed branches of the pigment cells of vertebrates. Immediate inspection suggests that crustacean chromatophores behave in the same way. Our most recent information about them (Perkins & Snook 1932) casts doubt on this. Outward streaming of pigment in crustacean chromatophores appears to accompany extrusion of cytoplasmic processes into collapsible pre-formed spaces. Extant observations on responses of isolated chromatophores both of vertebrates and of Crustacea refer almost exclusively to the action of light or of drugs. Most of what we know concerning chromatic behaviour is about the relation of stimulus to chromatic responses of intact animals. As such, it invites discussion under three main headings:

- (i) relevant stimuli and their appropriate receptors;
- (ii) integration of stimulus and response by nervous or endocrine co-ordination;
- (iii) its functional significance.

THE RECEPTIVE MECHANISM

In this context co-ordination implies the relation between a *specific stimulus*, a *specific receptor* and a *specific effector*. Consequently, thorough understanding of environmental agencies relevant to chromatic behaviour and of the receptor apparatus through which they operate is a necessary preliminary to fruitful investigation of nervous or endocrine control. It is also an essential prerequisite of experimental control, and supplies whatever clues we have for inquiry into the survival value of colour response—if any.

Electrical stimulation of specific areas (cloaca and roof of mouth) of some reptiles evokes generalized pallor, e.g. *Chameleo* (Hogben & Mirvish 1928) and *Phrynosoma* (Redfield 1918), or the reverse, e.g. *Anolis* (Kleinholz 1938b). Vigorous tactile stimulation of the cloaca of *Phrynosoma* and *Chameleo* has the same effect as faradization, and tactile manipulation of some teleosts modifies their normal photic responses. Otherwise response to electrical and tactile stimuli is not of widespread occurrence; and it has little to do with chromatic behaviour outside the laboratory. Inside the laboratory, as in nature, we may have to reckon with each or all of three significant external variables: light, heat and moisture. All animals known to display striking chromatic behaviour respond to photic stimulation of one sort or another. Most aquatic animals live in a relatively small temperature range, within which differences do not evoke striking pigmentary responses. It is roughly true to say that photic stimulation is the sole agency responsible for natural colour change of *aquatic* animals; but heat and desiccation may play a considerable part in the behaviour of *terrestrial* ones.

Most available data based on refined measurement of chromatic behaviour in relation to external agencies refer to the influence of light. Throughout a very wide range of species three types of photic response may exist with greater or less predominance of each one. One type, the *primary* or dermal reaction, is independent of the eyes and of the direction of incident illumination. The other two, which depend on the organ of vision, are the *ocular* or 'background' response to overhead illumination in light-absorbing (*black background*) and in light-scattering (*white background*) surroundings.

A stimulus which evokes dispersion of chromatophores of one type may evoke aggregation of chromatophores of another and *vice versa*. Thus the xanthophores of some vertebrates react in the opposite way to melanophores, and the leucophores of shrimps may react in the opposite way to xanthophores and erythrophores. For the sake of brevity, it will suffice to define the photic responses of vertebrates and of Crustacea in terms of the behaviour of those chromatophores which contribute most to the visible aspect of the organism, i.e. of the common reactions of the erythrophores and xanthophores of shrimps and that of the melanophores of other Crustacea and of vertebrates. With this restriction the *primary* (dermal) reaction is *sustained* dispersion of the erythrophores and xanthophores of shrimps and of the melanophores of vertebrates or of Crustacea other than Natantia. The white background (*tertiary ocular*) response involves their aggregation. The black background (*secondary ocular*) response is similar to the dermal reaction.

We can discriminate between dermal and ocular types of response by comparison of: (a) behaviour of normal and eyeless or blinded animals; (b) response at different intensities of illumination; (c) successive stages of chromatic behaviour under similar conditions. Thus the melanophores of *Xenopus laevis* (Amphibia), *Chameleo* pumulus* (Reptilia) and *Ligia oceanica* (Isopoda) are most aggregated in a dim light-scattering environment and most dispersed if the animals receive bright illumination from above in light-absorbing surroundings. After removal of the eyes (*Xenopus*) or after covering them with an opaque mask (*Ligia*, *Chameleo*), melanophores are relatively dispersed in bright light and relatively aggregated in complete darkness. If the eye is intact, transition from darkness to light-scattering surroundings in bright light involves an initial and transitory dispersion of pigment in the melanophores of *Chameleo*, of some Amphibia, e.g. *Amblystoma* (Laurens 1915, 1917), and in the erythrophores or xanthophores of some Crustacea, e.g. *Hippolyte* (Keeble & Gamble 1904).

Table 1 shows representative values of μ for the dermal melanophores of an amphibian, *Xenopus laevis*, and of the tail melanophores of an isopod, *Ligia oceanica*, (Smith 1938) in relevant situations.

Owing to its shape it is relatively easy to cover either the head alone or the rest of the body of a chameleon with black opaque cloth. Records of *Chameleo* (Zoond & Bokenham 1935) based on an arbitrary numerical scale of steps (from 1.0 to 5.0) referable to visible skin coloration then show the same contrast

* The Cape chameleon, now named *Lophosaura pumula*.

between dermal and visual stimulation (table 2). The dermal reaction like that of *Xenopus* and of *Ligia* has a higher threshold; and like that of *Amblystoma* and of *Xenopus* is much faster than the background responses. Such observations also suggest that rays of the red end of the spectrum are a relatively more efficient stimulus to the black background ocular response as also to that of *Xenopus* (Hogben & Slome 1936).

TABLE 1

	<i>Xenopus</i>		<i>Ligia</i>	
	bright light	dim light	bright light	dim light
A: normal animals:				
white background	2.0	1.5	1.7	1.4
black background	4.7	4.2	5.0	4.6
darkness		2.5		2.7
B: blinded animals:				
white background	3.1	2.7	4.2	3.9
black background	3.1	2.7	4.2	3.9
darkness		2.5		2.7

TABLE 2

	illumination by bright sunlight of			illumination by dim artificial light of		
	whole animal	skin only	eyes only	whole animal	skin only	eyes only
A: black background	4.7	4.3	3.7	4.9	2.0	4.3
B: white background:						
after 2 min.	3.5	5.0	1.8	4.0	2.5	3.1
after 20 min.	2.4	4.7	1.5	4.0	2.5	3.1

Since the dermal response is the only photic reaction of animals after removal of eyes, after blinding by covering the surface of the eye or after optic nerve section, it must be either: (a) an unco-ordinated response due to direct reaction of a pigmentary effector organ to the incident stimulus, or (b) a co-ordinated response involving some receptor organ other than the eyes. Reactions of the pigmentary effector organs in isolated skin, and response to change from light to darkness after destruction of the central nervous system or stoppage of the circulation, provide clear evidence that melanophores of Amphibia do in fact respond to light directly. So do those of some reptiles (Kleinholz 1938c).

The skin of *Chameleo rhodesiensis* will register an incident pattern of light and shade like a sheet of bromide paper. The possibility of evoking such purely local reaction suggests, but does not conclusively prove, that the primary, which is the predominant, photic response of such reptiles as the chameleon is autogenous; but evidence from other sources conclusively shows that the skin of cyclostomes (Young 1935a), and of Amphibia (Osborn 1938), contains photoreceptors which enter into certain types of locomotory behaviour. If melanophores have a direct

innervation (as is true of some reptiles and of Teleostei) we have therefore to reckon with the possibility that a localized primary reaction to light is a reflex phenomenon involving dermal photoreceptors and a sympathetic efferent path. In the absence of conclusive evidence, the existence of a nicely adjusted system of skin-autonomic reflexes capable of bringing about one-to-one local congruence of stimulus and response with the finesse of a photographic negative is not intrinsically likely.

The fact that pituitary removal abolishes the primary response of many vertebrates, including the lizard *Anolis carolinensis*, is not an insuperable objection to the view that it is a direct response to light. The presence of the hormone may sensitize the melanophores themselves to the action of light. That is to say, a minimal quantity of the relevant hormone in the circulation may be a necessary condition of melanophore dispersion. What tells most forcefully in favour of the view that the dermal response of reptiles depends on dermal photoreceptors is that the melanophores of a denervated area of the skin of the chameleon remain dispersed in darkness (Hogben & Mirvish 1928; Zoond & Eyre 1934).

In keeping with neither of the possibilities stated above is a suggestion that the *generalized* primary response of lampreys is a co-ordinated one for which the receptor is the pineal. Available data (Young 1935*b*) conclusively show that pinealectomy prevents the normal diurnal rhythm of colour change in *Ammocoetes*, but their correct interpretation is not equally clear. If the pineal is the photoreceptor, we might expect to produce the same effect by removal of the pineal or by withdrawal of light. This is not what happens. After pinealectomy melanophores of *Ammocoetes* remain permanently in the condition (dispersed) appropriate to illumination; but removal of the pineal complex does not abolish the primary response of the Urodele *Amblystoma* (Laurens 1916). Whether the responses of all animals which exhibit reactions to light stimuli without the intervention of the eyes are comparable is still uncertain. One fact which militates against generalization from the indisputable power of pigmentary effectors to react to light directly is the equally indisputable fact that a more or less subdued diurnal rhythm independent of temperature may persist in darkness, especially among Crustacea (Kleinholz 1937) and lampreys (Young 1935*b*). The nature of this persistent rhythm which keeps in step with the normal response to light and darkness is still obscure.

Exploration of the primary dermal reaction is a necessary preliminary to an understanding of the ocular or so-called background responses. At first sight, the pallor induced by exposure to light-scattering surroundings suggests the simple hypothesis that ocular stimulation brings into action a process capable of inhibiting the primary darkening of blinded animals exposed to light. Of itself this cannot explain two facts exhibited in tabular form above: (a) more intense illumination may produce more intense darkening of normal animals in light-absorbing (*black background*) surroundings; (b) blinded animals exposed to bright light in light-scattering (*white background*) surroundings are less dark than seeing animals in

bright light on a black background. The existence of the black background response calls for the additional hypothesis that ocular stimulation brings into action a mechanism which augments the primary response.

There is nothing inherently paradoxical in this supposition. Black background illumination involves stimulation of a restricted region of the eye, while illumination on a white background involves stimulation of the eye as a whole. This suggests: (a) that the black background reaction involves the stimulation of a specific receptive area, the *B* area, alone; (b) that an overriding reaction comes into play when light impinges on another set of receptors (the *W* area). We can put these assumptions to the test in various ways, more particularly: (a) by reversing the normal direction of incident illumination, (b) by covering the upper or lower part

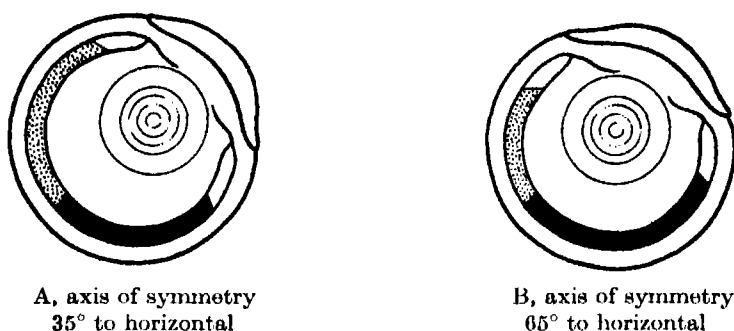


FIGURE 1. *B* and *W* areas of the eye of *Xenopus laevis*. These figures based on measurements made for the author by Miss M. I. Clarke show more precisely than the somewhat schematic figure previously published by Hogben & Slome (1936), the limits within which the effective *B* and *W* areas of the eye of *Xenopus* lie in transverse (A) and in longitudinal (sagittal) section (B). The lens of *Xenopus*, like that of a fish, has a refractive core which is almost exactly spherical. The region illuminated by all rays which traverse the lens from an extra-aqueous source is black. The region which can be illuminated only by rays reflected from sub-aqueous surfaces is dotted. The method of reconstruction is as explained by Hogben & Landgrebe (1940).

of the eye. With due regard to the orientation and structure of the eyes of different species, results of both tests conclusively demonstrate the localization of distinct receptive areas respectively concerned with the black and white background responses of Amphibia (Hogben & Slome 1936), of fishes (Hogben & Landgrebe 1940; Sumner 1933; Butcher 1938) and of Crustacea (Smith 1938).

The *W* area of the semi-terrestrial isopod *Ligia* is a latero-ventral group of ommatidia, the *B* area a dorsal group. Painting the lower half of the eye eliminates the white background response, and inferior illumination evokes it in light-absorbing as well as in light-scattering surroundings. The vertebrate eye is quite different. All light entering the eye of a vertebrate passes through the single lens, and, if the animal is aquatic, illumination from above in light-absorbing surroundings ordinarily restricts stimulation to rays falling inside a cone of which the half angle is the critical angle for air and water. The ventral area of the retina on which these

rays impinge delimits the effective *B* area. Accordingly, illumination from below in light-absorbing surroundings abolishes the black background response. The existence of the two areas must involve extrinsic differentiation of photoreceptors with reference to their central connexions. It may also involve intrinsic differentiation, i.e. of photoreceptors *inter se*. In support of this possibility there are indications of: (a) histological differentiation of the two areas of the teleostean eye (Butcher 1938; Neill 1942); (b) differential response of the *B* and *W* areas of the amphibian eye to different regions of the spectrum (Hogben & Slome 1936).

Analysis of the relevant areas of the eye of *Ligia oceanica* tallies with gross morphological differentiation of the eyes of some other arthropods (e.g. *Phronima*, *Gyrinus* and *Bibio*), and recalls Mast's analysis of phototactic behaviour among syrphids. Capacity to execute background response and phototaxis both depend on regional orientation of photoreceptive elements to incident illumination. Equally sharp delimitation of a *W* or a *B* area in the sessile eye of an isopod, an amphipod or an insect is possible because each ommatidium admits a minute sheaf of rays. Clearly, the same conditions of regionalization do not hold for the eye of a terrestrial vertebrate. In aquatic vertebrates regionalization of the stimulus is guaranteed by extrinsic circumstances. When the aqueous environment does not scatter light, the spread of the most oblique rays from an extra-aqueous source is limited by the critical angle. There can therefore be a relatively small and sharply defined *B* area. In light-absorbing surroundings the horizon sets the only limit to the obliquity of rays impinging on the retina of a land vertebrate. So there cannot be a small sharply defined *B* area, and the effective *W* area must be restricted to the region stimulated by rays reflected upwards. In the eye of a fish (Hogben & Landgrebe 1940) this region is neutral. That is to say, inferior illumination in light-absorbing surroundings abolishes the white background as well as the black background response. The *W* area of the eye of a typical teleost is a relatively small region in the neighbourhood of the optic nerve.

The fact that background responses depend on a highly specialized adjustment involving regional differentiation of receptive elements, orientation of the eye with reference to the head, and also its freedom of movement, throws some light on the vagaries of photic reaction among closely allied species. It also helps us to understand why a comparatively small number of species among the many with a well-developed pigmentary effector system show striking changes of colour in relation to normal variations of incident illumination. Any considerable variation of the orientation of the eye or of its mobility will put the mechanism of background adjustment out of gear, and it may do so in either of two ways. The overriding *W* mechanism may come into play in nearly all circumstances when the animal is exposed to light. Unless the primary reaction has a lower threshold to stimulation than the ocular responses it will then be impossible to detect a primary response without recourse to observations on (a) behaviour of blinded animals, or (b) the time sequence of chromatic change. Alternatively, stimulation of the effective *W* area may never come into play in normal circumstances. The reaction of the

animal to light will then be exactly the same as that of blinded individuals of species, e.g. the chameleon, with a strong primary response.

There is thus room for great diversity of chromatic behaviour in response to light by the interplay or individual predominance of three fundamental types of photic response known to coexist in one and the same species of groups so diverse as Isopoda (*Ligia*), Decapoda (*Hippolyte*), fishes (*Raia*), Amphibia (*Xenopus*, *Amblystoma*), Reptilia (*Anolis*, *Phrynosoma*, *Chameleo*). Table 3 sets forth the several processes which light initiates, and suggests one way of avoiding misunderstandings now prevalent. It would greatly clarify disputed issues if all workers on chromatic behaviour agreed to restrict the use of:

(a) The term *white background response* to what happens when we transfer the intact animal from superior illumination in light-absorbing to superior illumination in light-scattering surroundings.

(b) The term *black background response* to what happens when we transfer intact animals from darkness to light-absorbing surroundings illuminated from above by light of intensity insufficient to evoke a response of the blinded animal.

TABLE 3. THE THREE COMPONENTS OF PHOTIC RESPONSE

In this table to illustrate the different processes which light brings into play:

I = *primary* (dermal) response, dispersion of melanophores.

II = *black background* or *secondary* (ocular) response, dispersion of melanophores.

III = *white background* or *tertiary* (ocular) response, aggregation of melanophores.

'Dim' refers to light of intensity below the threshold of stimulation for I but above threshold for II and III, 'bright' of intensity above threshold for I, II and III.

illumination (from above)			dim		bright			
surroundings	...	...	light- absorbing (black background)	light- scattering (white background)	light- absorbing (black background)	light- scattering (white background)		
response :								
(a) normal animal			II	II and III	I and II	I and II and III		
(b) blinded animal			0	0	I	I		

Information concerning the relation of chromatic behaviour to heat and to cold or to drought and to moisture is restricted to a few terrestrial forms. To the best of our knowledge temperature is a major variable of colour response only among reptiles; but precise information is meagre (Parker 1903, 1938). Few investigations, and none of them based on actual measurement of relative humidity, deal specifically with the influence of moisture and of drought on the chromatic behaviour of terrestrial animals. Some Amphibia and phasmids respond noticeably to moist and to dry air at fixed temperature and standard illumination. Among phasmids (Giersberg 1930) a generalized response occurs as the result of exposing only the hind end of the body to one or the other. Thus a form of irritability restricted to the pharyngeal region of man may be localized in other parts of the animal body. A somewhat

archaic observation on toe-thirst (Biedermann 1892) is noteworthy in this connexion. According to an unconfirmed statement the tree frog, *Hyla arborea*, does not respond by colour change to moisture after narcotization of the pads at the end of the digits of the hind legs.

What little we do know about the role of temperature and relative humidity as agents of chromatic behaviour refers largely to terrestrial or semi-terrestrial vertebrates and to insects. Tentatively, we might perhaps sum up our scanty information about the former in the two statements: (a) high temperature and low humidity beyond a certain limit either abolish the primary response to light or inhibit the effect of stimulation on the *B* area of the eye; (b) action of humidity almost certainly, and of temperature possibly, involves skin receptors. For the present we have therefore to limit profitable discussion of the problem of co-ordination to the more widely spread reactions evoked by light.

CO-ORDINATION OF COLOUR RESPONSE

Early investigators of chromatic behaviour, such as Lister, Pouchet and Biedermann, postulated direct innervation of the pigmentary effector system of vertebrates and of Crustacea as the basis of co-ordinated colour response to localized stimulation. To-day there is general agreement about the important role which internal secretion plays. Though there is no unanimity concerning the possibility of harmonizing results of investigations on closely allied species within the framework of any single comprehensive hypothesis, there is little doubt about three general conclusions:

(i) Maximal dispersion of the melanophores of cyclostomes, Amphibia, elasmobranchs, of some Teleostei (e.g. *Anguilla*, *Ameiurus*) and of some reptiles (e.g. *Anolis*) is contingent on the presence in the blood of a hormone secreted by the *pars intermedia* of the pituitary gland.

(ii) Maximal dispersion of melanophores and erythrophores of crabs and maximal aggregation of chromatophores containing the yellow or red pigments of shrimps is contingent on the presence in the blood of a hormone produced in the eye stalk, probably by the *sinus gland* (Hanström 1937).

(iii) While there is no conclusive experimental or histological evidence for the direct innervation of the pigmentary effector system of Crustacea, cyclostomes, Amphibia or elasmobranchs, evidence of both kinds points to the conclusion that the melanophores of Teleostei and of some reptiles are under peripheral (sympathetic) control of nerves. Suitable stimulation of these fibres brings about aggregation of melanophore pigment.

Of (iii) it is enough to say that segmental effects of nerve stimulation after stoppage of the circulation complement the observation that nerve section leads to local abolition of certain responses (Hogben & Mirvish 1928; Parker & Rosenblueth 1941; Parker 1941). Evidence for (i) and (ii) is based on effects of removal of the organs mentioned, on complementary effects of their extracts, and on

assay of the concentration of excitant substances with corresponding properties in the circulation.

Thus hypophysectomy abolishes the black background response of *Anura* (Allen 1917; Smith 1920; Hogben 1923; Krogh 1922; Hogben & Winton 1923; Houssay & Ungar 1925; Hogben & Crew 1923), of elasmobranchs (Lundstrom & Bard 1932; Hogben 1936; Waring 1938), of the lizard, *Anolis carolensis* (Kleinholz 1938*b*), and (with the qualification stated below) of the eel, *Anguilla anguilla* (Waring 1940). It also abolishes the normal dispersion of the melanophores of *Lampetra* (Young 1935), larva or adult, as also of *Urodeles* (Hogben 1924) in light. Implantation of the *pars intermedia* of the pituitary into anuran tadpoles (Swingle 1921) evokes melanophore dispersion in individuals which would otherwise be pale. Extracts of the *pars intermedia* evoke melanophore dispersion in denervated areas of some species of *Teleostei* (Kleinholz 1935), as also in isolated pieces of skin of *Anura* (Hogben & Winton 1922*c*, 1922*b*) and of *Anolis* (Kleinholz 1938*c*). Injection of such extracts restores melanophore dispersion after hypophysectomy of *Anura* (Hogben & Winton 1923; Hogben & Gordon 1930; Houssay & Ungar 1925), of *Urodela* (Hogben & Crew 1923; Hogben 1924) and of *Elasmobranchs* (Lundstrom & Bard 1932; Hogben 1936; Laurens 1915, 1917).

The active *B* substance of the *pars intermedia* is present in the pituitary of all classes of vertebrates (Hogben & de Beer 1925). A method of assay based on the response of *Xenopus*, preferably hypophysectomized (Hogben & Gordon 1930; Landgrebe & Waring 1941), yields results of reliability comparable to assay of other posterior lobe autocoids, with whose properties it is therefore possible to compare its own. It shares some characteristics of the oxytocic substance of Dale and of the pressor substance of Schafer, being dialysable, resistant to boiling in slightly acid solution, and readily destroyed by trypsin and strong oxidizing agents (Hogben & Winton 1922*a*, *c*). Unlike both, it resists treatment with cold strong alkali (Hogben & Gordon 1930; Landgrebe & Waring 1941). It is possible to separate it from them by various more or less drastic methods of fractionation but also by adsorption to animal charcoal (Landgrebe & Waring 1941). Clearly therefore it is a distinct entity.

The pituitary of frogs (Hogben & Winton 1922*a*) stores far more of the *B* substance than the threshold dose which overrides the white background response; but the amount in the gland of the eel (Waring & Landgrebe 1941) is about one-tenth as great as the total quantity in its circulation when kept in light on a black background. Hypophysectomy produces permanent pallor of some species of *Amphibia*, e.g. *Amblystoma tigrinum* (Hogben 1924), and of elasmobranchs, e.g. *Raia clavata* (Hogben 1936), in spite of the fact that the pigment in their melanophores is almost completely dispersed in darkness and fully dispersed when the intact animal remains in light on a white background. The concentration of *B* in the blood of such species never falls to the level which permits aggregation, and failure to respond to different conditions of illumination cannot be due to any functional defect of the pigmentary effector organs themselves.

Endocrine control of crustacean colour change is suggested by negative results of nerve section (Koller 1927; Perkins 1928), and attested by effects of injecting the blood of dark into pale animals or vice versa (Koller 1927; Kleinholz 1937). Totally different results obtained on the one hand by cauterizing or by covering the ommatidia of a decapod with an opaque mask and on the other by total extirpation of the eye-stalk suggest that: (a) the eye-stalk is the seat of liberation of a hormone; (b) the production of it is under reflex retinal control. Thus blinding by removal of the distal end of the eye-stalks of *Uca pugilator*, which is typical of Brachyura, has no appreciable effect in light, but complete extirpation of both eye-stalks leads to permanent aggregation of melanophores and erythrophores (Abramowitz 1937; Carlson 1936). Cauterization of the retinal part of the eyes of *Palaemonetes*, which is typical of decapods other than Brachyura, leads to medium dispersion of the erythrophores and xanthophores, but complete ablation of both stalks produces extreme dispersion of the red and yellow pigments in light on a white background (Parker 1941). Injection of aqueous extracts of the eye-stalks produces complementary effects. That is to say: (a) it evokes dispersion of the melanophores or erythrophores of *Uca pugilator* as of other Brachyura so far studied; (b) it leads to aggregation of the red and yellow pigments of *Palaemonetes*, and in general of decapods other than Brachyura.

Extracts of eye-stalks from shrimps evoke dispersion of the melanophores and erythrophores of crabs, and extracts of eye-stalks of crabs evoke aggregation of the red and yellow pigments of shrimps. Seemingly, one and the same substance therefore controls both: (a) the black background response of Brachyura (and possibly of Stomatopoda); (b) the white background response of Astacura, of Anomura and of Natantia. The same hormone also appears to be present in the heads of Cumacea, amphipods and isopods (Ståhl 1938); and head extracts of the isopod *Ligia baudiniana* evoke aggregation of melanophores of the same species (Kleinholz 1937). The eye-stalk hormone of Decapoda is a dialysable, water-soluble substance which resists boiling in presence of dilute acid or alkali and the action of proteolytic enzymes. It is not soluble in organic solvents such as ether and chloroform. Judged by the quantity of tissue required for extraction of the threshold dosage, eye-stalk extracts are highly specific. In each eye-stalk of *Uca* there is sufficient of the active substance to evoke in one blinded individual melanophore dispersion which does not completely subside till five hours have lapsed (Abramowitz 1937a).

Though there is general agreement with reference to the three issues specified above, three others are still open to controversy. These are:

(i) Whether co-ordination of chromatic behaviour of some Teleostei and of some reptiles is wholly independent of pituitary secretion.

(ii) Whether reflex activation or inhibition of the secretion activity of the pars intermedia is the sole agency of co-ordination in other species of vertebrates.

(iii) Whether reflex activation or inhibition of the sinus gland is the sole agency for co-ordination in Crustacea.

It will be easier to assess what essential similarities underlie the diverse phenomena characteristic of species within the same or closely related groups when we have more relevant information. Meanwhile diversity of views expressed by different investigators with reference to the above arise less from disagreement concerning experimental data on record than from different views about correct interpretation. In connexion with disagreement of this sort two issues call for comment.

One of them, specially relevant to the control of colour response among vertebrates, involves a principle pointed out long ago by Langley. Peripheral nerves generally carry vasomotor neurones. Section and stimulation of nerves will therefore bring about increase or decrease of blood flow through regions they supply. If the blood contains a hormone to which pigmentary effector organs respond, the result may be increase or decrease of the quantity of an excitant substance in their vicinity. Where we have reason to suspect the existence of such a hormone in the blood, we should therefore hesitate to draw conclusions from local effects due to local section, anaesthetization or stimulation of nerve trunks in animals *with an intact circulation*. Experimental evidence for direct nervous control of the pigmentary effector system is conclusive when it is possible (Hogben & Mirvish 1928) to demonstrate such effects in isolated segments of the body or after stopping the circulation of the animal as a whole.

A second interpretative issue which calls for comment is implicit in what has gone before. Thorough study of external conditions which evoke a particular type of response is an indispensable prerequisite for experimental analysis of the mechanism of control; but many recorded observations on the effects of nerve stimulation or section, of removal of ductless glands or injection of endocrine preparations refer to species of whose normal behaviour we have insufficient knowledge. Unnecessary confusion therefore results from discussing what controls the 'light phase' and the 'dark phase', or what controls 'darkening' and 'blanching'. Each epithet (table 3) can stand for several different physiological phenomena. Pallor may be due to absence of light, so that neither a primary dermal reaction, nor the effects of ocular stimulation of the *B* area, come into play. It may also be due to the mechanism of co-ordination brought into play by the action of light on the *W* area of the eye, overriding both the primary reaction and the process of co-ordination due to the stimulation of the *B* area. Likewise, the 'dark phase' may be: (a) a manifestation of the primary reaction alone, as in a blinded animal; (b) a result of exclusive stimulation of the *B* area of the eye; (c) an outcome of both processes.

Controversy which has arisen in connexion with the role of the pituitary in the chromatic behaviour of Teleostei illustrates several misunderstandings of this sort. So far as we know, most teleostean fish are pale when kept in the dark, and few have a detectable primary response. If such fish show background responses, we have to distinguish between at least two kinds of pallor, and hence with two different problems of co-ordination. Darkness pallor is what happens when the

melanophores are affected neither by the mechanism of co-ordination brought into play by photic stimulation of the *B* area of the eye nor by the mechanism of co-ordination brought into play by stimulation of the *W* area. White background pallor happens when stimulation of the *W* area brings into play a co-ordinating process which *overrides* the effect of stimulating the *B* area. If injection of pituitary extracts containing the *B* substance evoked no reaction from melanophores of a pale fish kept in darkness, we should have to conclude that the pituitary is not concerned with the black background response; but admitted failure (Matthews 1931) to evoke this reaction from fish kept on a white background is irrelevant.

Whatever is responsible for the white background response antagonizes the result of stimulating the *B* area and of bringing into play the process of co-ordination involved in the black background response. So we are not forced to expect any positive result from introducing into the circulation of an animal subject to photic stimulation of the *W* area the putative agent of the black background response. In other words, failure to elicit reaction of fish made pale by exposure to a white background cannot dispose of the possibility that liberation of the *B* substance of the *pars intermedia* is responsible for the black background reaction of teleosts as of other vertebrates. *B*-containing extracts which do not evoke darkening of pale fish on a white background, do in fact produce dispersion of melanophores in a denervated area of the same species (Kleinholz 1935).

By the same token, the mere fact that hypophysectomy prevents darkening of a fish after transference from white to black background in light does not necessarily prove that reflex activation of the pituitary governs the black background response. Before we can draw this conclusion with assurance we need to know something about the behaviour of the *blinded* animal. Darkening after transference from darkness to black background in strong light may be due to the primary or to the secondary ocular reaction or to the outcome of each reinforcing the other. Darkening after transference from white to black background in strong light may be due to cessation of the tertiary ocular reaction which overrides either or both of the preceding. When transferred from darkness to a black background in strong light or from an illuminated white to an illuminated black background and vice versa, an intact fish with a strong primary (dermal) but with no secondary (ocular) reaction would behave exactly like a fish with a strong secondary but no primary reaction. So far we know little about the effects of hypophysectomy after removal of the eyes or section of the optic nerves of Teleostei, and not enough about effects of blinding of otherwise normal individuals in this group.

Thus some controversial issues might well resolve themselves if the same sources of relevant information were available for different species under discussion. Analogous remarks reply to Crustacea, as to Teleostei. It would be easier to assess similarities and differences among related species of decapods, if we had more available data concerning chromatic behaviour of animals deprived of ocular stimulation by different methods. The results of painting the eye of a vertebrate, of cutting the optic nerves and of removing the eye as a whole are the same; and

the last-named procedure is the most convenient way of isolating the primary response to light. Since the sinus gland of decapod Crustacea is in the eye-stalk itself, removal of the latter is not an operation comparable with section of the optic nerves or with putting an opaque cover over the outside of the eye. Hence it does not necessarily give any indication of the presence or absence of a detectable primary response. It may be that some of the apparent anomalies of crustacean colour response will disappear when we have more information about the distinctive effects of eye removal and of blinding without removal of the sinus gland.

Among fruitful clues which merit more attention are those we can get from quantitative study of time relations. Though all investigators of chromatic behaviour now recognize the existence of relatively *slow* responses co-ordinated by reflex internal secretion and relatively *speedy* ones co-ordinated by direct innervation, few workers have given sufficient attention to the information which quantitative study of time relations can furnish. Study of the time relations of chromatic behaviour is illuminating in several ways. We have already seen that it may help us to discriminate between dermal and ocular components of photic response, because the former may involve a faster reaction than the latter. In the same way, we can distinguish between pallor due to stimulation of the *W* area of the retina of a fish exposed to superior illumination in light-scattering surroundings and pallor due to lack of stimulation of either the *B* area or the *W* area of the retina of a fish exposed to *inferior* illumination in light-absorbing surroundings (Hogben & Landgrebe 1940). The study of time relations can also throw light on (a) whether direct innervation of the pigmentary effector system contributes to co-ordinated responses; (b) whether co-ordination is brought about by secretion or excretion of a single, in contradistinction to more than one, hormone.

The period which elapses between onset of stimulation and completion of any co-ordinated response involves the latent period of the receptor, the time necessary for the transmission of the disturbance through the nervous system or blood and the reaction time of the effector. We can get information about the reaction time of pigmentary effector organs from their responses to chemical or electrical stimuli (Neill 1940; Waring & Landgrebe 1941). It varies from somewhat less than a minute (some Teleostei) to half an hour (*Anguilla*) or even 100 min. (*Xenopus*). In general, the time lag involved in the transmission of a stimulus to a pigmentary effector organ by a purely nervous path is therefore trivial compared with the reaction time of the end-organ. If the time which elapses between onset of stimulation and a new equilibrium state of the intact animal considerably exceeds the reaction time of the pigmentary effector, we may safely conclude that any generalized response to a local stimulus is, at least predominantly, the result of reflex activation or reflex inhibition of an endocrine organ.

That fluctuating secretion of one and the same endocrine organ can furnish a sufficient basis for the control of co-ordinated colour response is an hypothesis which carries additional implications. For illustrative purposes, let us suppose that the equilibrium values of μ for a vertebrate with no detectable primary

reaction are respectively 2.5 and 1.5 in darkness and in light-scattering surroundings with superior illumination. The assumption that photic response is controlled by a single melanophore-dispersing hormone (*B* substance) means that the blood of an animal equilibrated on a white background contains less than the blood of an animal equilibrated in darkness; and if the value of μ for the black background reaction is 4.5 the blood of an animal kept in light on a black background should contain more than that of either. Transition from equilibrium on a black background to equilibrium on a white background in light should therefore involve the *elimination* of a greater quantity of the melanophore-dispersing hormone than transition from equilibrium in light on a black background to equilibrium in darkness or from equilibrium in darkness to equilibrium in light of a white background.

We should therefore expect that the time interval of equilibration (b^Tw) for transference from black to white background in light would be greater than the equilibrium interval (b^Td) for transference from black background in light to darkness, or the interval (d^Tw) of equilibration for transference from darkness to a white background in light. If we find that the reverse is true, i.e. $b^Tw < b^Td$ or $b^Tw < d^Tw$, we have strong reasons for rejecting the single-hormone hypothesis. Analogous remarks apply to reverse changes of external conditions. These involve the *liberation* of more or less *B* substance, if the same hypothesis is correct. If w^Tb stands for the equilibrium interval for transition from white to black background in light, and the meaning of w^Td , d^Tb is consistent with that of preceding symbols, the expectation is that $w^Tb > d^Tb$ or w^Td . *Mutatis mutandis*, the same test is applicable to the adequacy of a hypothesis which postulates reflex liberation or disappearance of a single hormone evoking aggregation of the melanophores, as does the sinus gland hormone of shrimps. Table 4 (Smith 1938) shows the results of applying this method of analysis to the background responses of the Isopod *Ligia*.

equilibrium interval	approximate time (hr.)	approximate range of μ	approximate μ shift
b^Tw	2	4.5-1.5	3.0
b^Td	20	4.5-3.0	1.5
d^Tw	3	3.0-1.5	1.5
w^Tb	6	1.5-4.5	3.0
w^Td	10	1.5-3.0	1.5
d^Tb	4	3.0-4.5	1.5

Inspection of these figures shows that a relatively small shift of μ in either direction requires a relatively long time. This is quite intelligible if we interpret the speed of crustacean background responses in terms of competition between rates of accumulation or elimination of two antagonistic hormones. It is not consonant with a one-hormone hypothesis; and another line of attack (Kleinholz 1938a) based on comparison of the threshold concentration of eye-stalk hormone for different classes of pigmentary effectors bears out this conclusion. Retinal pigment cells of prawns, like their integumentary erythrophores, react to the

eye-stalk hormone by aggregation. The latter respond in the opposite way to white and black background illumination; the former undergo aggregation in light on a black as well as on a white background. If a one-hormone hypothesis gives a correct account of the co-ordination involved in the responses of both types of pigmentary effector organs, the retinal pigment cells should therefore undergo aggregation at a lower threshold concentration than the integumentary erythrophores. Injection experiments show that this is not so, and the hypothesis that crustacean chromatic behaviour is controlled exclusively by the liberation or subsidence of the sinus gland hormone alone is inadequate. There is some evidence for the production of an antagonistic hormone by the rostral gland (Koller 1929) of decapods.

Analysis of the time relations of chromatic behaviour of *Xenopus laevis*, of the eel (*Anguilla anguilla*) and of the dogfish (*Scyllium canicula*) lead to analogous conclusions (Hogben & Slome 1931; Neill 1940; Waring, Landgrebe & Bruce 1942). The background responses of all three species are of the slow type pointing to reflex control of an endocrine organ; and the relation of equilibrium interval to μ shift does not harmonize with the simple hypothesis that the two types of response to retinal stimulation are brought about respectively by activation or inhibition of the pars intermedia. Evidence from a variety of sources casts doubt on the possibility that such a hypothesis is adequate for any vertebrate. One is the surprising results of determining threshold dosage of pressor-free extracts containing the *B* substance when injected into normal and hypophysectomized animals. The one-hormone hypothesis requires that animals would be equally tolerant* after removal of the pars intermedia alone or of the whole pituitary gland, and less tolerant with the gland intact. Experiments on *Xenopus laevis* (Hogben & Gordon 1930; Hogben & Slome 1931, 1936) and on the elasmobranch, *Scyllium canicula* (Waring 1938), show that this is not so. Removal of the whole pituitary decreases tolerance; and removal of the pars intermedia alone increases it.

The only straightforward explanation of this circumstance, now abundantly substantiated, is that some part of the pituitary gland other than the pars intermedia (i.e. the pars anterior or pars tuberalis) either (a) produces something (*W* substance) which has the opposite effect to the *B* substance secreted by the pars intermedia itself; or (b) controls the activity of another organ which itself produces a hormone antagonistic to the *B* substance. Either view implies that removal of the whole pituitary need not lower the melanophore index as much as removal of the pars intermedia alone. This is true of *Xenopus* (Hogben & Slome 1931, 1936). In fact, removal of the whole pituitary need not result in maximum pallor at all. Removal of the whole pituitary of the eel (Waring 1940) does not lower the melanophore index below 3.5, the equilibrium value for darkness. If withdrawal of the *B* substance were solely responsible for the protracted white background response total hypophysectomy should lower μ to 1.5.

Evidence of yet another type points to the same general conclusion. The

* High tolerance implies a high threshold dosage.

immediate sequence of events after hypophysectomy depends on whether an eel (Waring 1940) has been kept for a long while in light on a black background before the operation or has been kept just long enough to darken after protracted exposure to light on a white background. The two-hormone hypothesis prescribes that the blood of an eel of the latter type should have (a) insufficient of the *W* substance to lower the melanophore index appreciably; (b) more of it than the blood of an eel kept sufficiently long on a black background to permit further subsidence of the concentration of *W* in the blood; (c) therefore more of it than the blood of an eel kept in darkness. Since the time relations of the background response point to the conclusion that the concentration of *W* subsides more slowly than that of *B*, the ratio of *W* to *B* in the blood of such eels should rise immediately after hypophysectomy above its level in normal animals kept in darkness. The melanophore index should therefore fall below the corresponding equilibrium value of the hypophysectomized eel before reaching it. This is exactly what happens. After hypophysectomy the value of μ falls steeply and continuously to equilibrium (3.5) if eels have been long on a black background in light. If removal of the pituitary takes place when eels have had just long enough to darken after previously prolonged exposure to white background in light, the value of μ falls slowly to about 2.0 and gradually rises to 3.5 after a period about five times as long as the equilibrium interval for hypophysectomy after prolonged black background exposure.

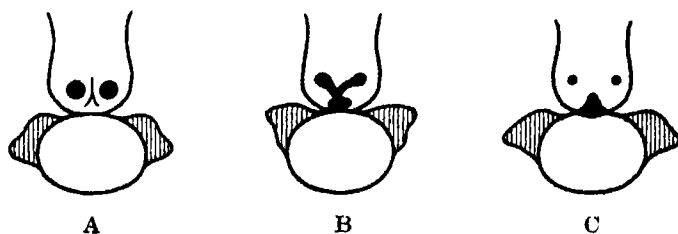


FIGURE 2. Diagrammatic representation of different types of Anuran pituitaries. (Pars intermedia shaded; pars anterior white.) A. *Adult Rana*: no median process, pars tuberalis represented by two glandular discs. B. *Bufo* at metamorphosis: pars tuberalis represented by bifid anterior process of the pars anterior. C. *Xenopus*: median anterior process of basophil cells and two inconspicuous discs of small cells apparently non-glandular.

In keeping with the results of so diverse methods of attack upon the problem, removal of the anterior lobe of *Xenopus* and of *Anguilla* (Hogben & Slome 1931, 1936; Waring 1940) abolishes the white background response. This is not true of the frog, *Rana temporaria* (Hogben & Winton 1923); but cauterization of the *pars tuberalis* of *Rana pipiens* (Steggerda & Soderwall 1939) has an effect analogous to that of removal of the anterior lobe from *Xenopus*. That such treatment abolishes the white background response might merely be due to damage of nerve tracts involved in reflex activation of the intermediate lobe (Giusti & Houssay 1924; Houssay *et al.* 1935); but the possibility of doing so does not dispose of the effect which cauterization of the *pars tuberalis* has on tolerance to *B*-containing extracts. Cauterization

of the pars tuberalis of *Rana pipiens* (Steggerda & Soderwall 1939), together with removal of the anterior and posterior lobes, makes the frog less tolerant to the *B* substance than removal of both the anterior and the posterior lobe without damage to the pars tuberalis itself.

Possibly the different effects of removing the anterior lobe from *Xenopus* and from frogs has its explanation in structural differences. Unlike that of *Rana* the anterior lobe of *Xenopus* has an anterior process (Hogben & Slome 1931; Atwell 1941), the cells of which are different from those of the rest of the *pars anterior*. The *pars tuberalis* of toads develops as a median forward process of the anterior lobe bifurcating at its extremity. At metamorphosis the distal branches separate as two plaques which come to lie on either side of the floor of the infundibulum. The neck itself usually disappears. Apparently, it persists in *Xenopus* as the 'anterior process' so characteristic of this species. The paired disks corresponding to the pars tuberalis of *Rana* are inconspicuous (Atwell 1941) and 'give slight indication of glandular capabilities'. Accordingly, we may regard the *anterior process* of *Xenopus* as the functional equivalent of the paired *pars tuberalis* of *Rana*.

One account of a pituitary preparation which evokes aggregation of Amphibian melanophores is on record (Geiringer 1935). There seems to be a dioxyphenyl compound, with properties like those of adrenaline, in the anterior lobe of the ox. Failure to prepare extracts of the amphibian pituitary capable of evoking pallor is not surprising in the light of the time relations of the white background response. Its protracted character suggests that the endocrine organ responsible for liberating the *W* substance has no storage capacity, but liberates its secretion *pari passu* with its manufacture. On the other hand, difficulty may arise from the fact that the pituitary plays an indirect role in relation to the white background response by controlling the functional activity of another ductless gland.

A suggestive, but as yet unconfirmed, observation (Stoppani 1941) relevant to this possibility is the assertion that removal of the suprarenals prevents *aclaramiento* (presumably the *W* response) of *Bufo arenarum* in response to photic stimulation through the eyes. It would be premature to draw the most obvious conclusion from this experience. The suprarenal of Amphibia is imbedded on the surface of the kidney, and its excision without damage to the latter is extremely difficult, if indeed practicable. Possibly, a result of this kind might therefore follow interference with the normal activity of the kidney and hence with the rate at which the *B* substance normally accumulates. If we take the effect of the operation at its face value, we are left with two possibilities. The hormone concerned with the white background response of Amphibia and of elasmobranchs may be adrenaline or it may be a product of the cortical cells. The reaction of vertebrate melanophores to adrenaline does not favour the view that adrenaline is the agent. Though it is true that melanophores of many species react to adrenaline by aggregation, the threshold concentration necessary to evoke the reaction is often vastly in excess of the quantity which can be present in the blood, and the slight reaction of many

Amphibia to an effective dose might well be an indirect consequence of vasoconstriction.

From an evolutionary point of view, the chromatic behaviour of the eel, and the information we have about its control, are of special interest. Everything stated about *Anguilla* in preceding paragraphs falls into line with what we know about Amphibia and about elasmobranchs; but other observations point to the emergence of a new mechanism of co-ordination. Hypophysectomized eels give a slight response to background; and stimulation of isolated segments shows that its dermal melanophores have direct innervation. On the other hand, the time relations of the two responses before and after removal of the pituitary point to the conclusion that nervous control as yet plays no appreciable role in normal chromatic behaviour.

Between that of the eel at one extreme and of a teleost such as *Lebistes* at the other (Neill 1940) speed of chromatic response has many grades. When transferred from white to black background in light some teleosts can change colour more or less completely in a few seconds; and comparison of the equilibrium interval for transference from black to white background in light with the reaction time of the melanophores themselves suggests that the *white background response* is brought about mainly, if not exclusively, by direct nervous control of the pigmentary effector system. In all probability, direct nervous control is mainly responsible for the *initial* phase of the reaction of those fishes which respond more slowly; but it cannot provide a satisfactory clue to what happens in the later stages. The stickleback *Gasterosteus*, which is visibly pale within 10 min. after transference to a white from a black container in light, does not attain maximal melanophore aggregation till the lapse of about 30 min. (Hogben & Landgrebe 1940). The reaction time of the melanophores of *Gasterosteus* is about 6 min. So it is not likely that direct nervous control accounts for the completion of the change. Since the *B* area is still subject to stimulation (table 3) when an animal is on a white background, speed of colour change in response to change of background in light gives us no clue to the control of the black background response. The relevant time relations are those of transference from black background in light to darkness or vice versa. The equilibrium interval for transference of *Gasterosteus* from black background in light to complete darkness is about a day. The pallor that supervenes in darkness is apparently due to dissipation of the hormone responsible for the black background response. Presumably this is the secretion of the pars intermedia.

Evidence from different sources leads us to believe that nervous control may partly or wholly take the place of reflex endocrine co-ordination of the black background response of other Teleostei. Histological (Ballowitz 1893) and experimental (Parker & Rosenblueth 1941) evidence provides strong support for the view that teleostean melanophores may have a double innervation. Recorded effects of removing the pituitary from *Fundulus* (Matthews 1933; Abramowitz 1937c) point to two conclusions: (a) that it does not completely abolish response of the intact fish when transferred from white to black background; (b) that it does abolish response of *denervated* melanophores to the same treatment. Hypophysectomy of the catfish, *Ameiurus*

nebulosus (Osborn 1938), reduces pigment dispersion after transference from white to black background response by about 50 % (micrometer assay). Unfortunately, we know very little about the conditions relevant to the manifestation of a *primary* response in Teleostei. What legitimate conclusions we can draw from such experiments will therefore be more clear when we have information about the effects of hypophysectomy on *blinded* as well as upon normal individuals. Meanwhile it is probable that nervous co-ordination has partly usurped the more archaic type of black-background control common to Amphibia, elasmobranchs and to the eel itself. It is possible that it has completely superseded reflex stimulation of the *pars intermedia* of a few species.

FUNCTIONAL SIGNIFICANCE OF CHROMATIC BEHAVIOUR

Widespread existence of chromatic behaviour in the animal kingdom, diversity of external stimuli which bring it into play and the several types of control associated specifically, as it seems, with its manifestations, tempt us to conclude that it must have some survival value. So far we have no conclusive proof that it has; and what survival value it may have is susceptible to direct investigation of a limited range of species.

Chromatic behaviour of terrestrial animals invites the exploration of a possibility which is accessible to experiment in the laboratory. Land animals which exhibit chromatic behaviour are poikilothermous. Presumably the state of the pigmentary effector system affects absorption of radiant heat when the body temperature is below, and loss of heat by radiation when it is above, that of the surroundings. The fact that some animals react to heat by pallor therefore suggests that chromatic behaviour may play a part in the control of body heat. Thermocouple measurements of skin temperature, and observations on the time of survival, of pale and dark animals exposed to the same source of radiant heat sufficiently long to raise the temperature of the body to the upper thermal death point would settle whether the suggestion is correct.

Most animals which exhibit chromatic behaviour are aquatic, and the responses of aquatic animals are predominantly photic. Conceivably, the background response may make such species less conspicuous to their natural enemies, to their competitors or to their prey; but more relevant ecological information than we now have at our disposal is a necessary prerequisite of laboratory investigations designed to explore this possibility. Meanwhile, experimental biologists are no longer disposed to speculate at large about protective or warning coloration without the precaution of ascertaining what range of stimuli determine locomotory behaviour* involved in the feeding habits of a predator or the avoiding reactions of its prey.

* Popham, E. J., *Proc. Zool. Soc. A*, 111, 135 (1942).

GENERAL CONCLUSIONS

With due regard to the need for more information and to uncertainties arising from the description of totally different phenomena as pallor or otherwise, the following provisional scheme is compatible with well-established available evidence. It has at least one merit. It points the way to inquiries which would clarify controversial issues, whether they wholly sustain it or force us to modify it.

(i) Phototactic responses of vertebrates and of Crustacea involve three components. Any one of them may predominate. Any one of them may be negligible. All of them may coexist in one and the same species, contributing more or less to the ordinary rhythm of colour change. The three components are as follows:

(a) A *primary dermal* reaction, which is independent of the eyes. This is probably the most archaic, since it is the only type of phototactic response among Hirudinea which possess melanophores. It is also the characteristic light reaction of such reptiles as *Chameleo* or *Phrynosoma*, possibly also of some elasmobranchs.

(b) A *secondary ocular* reaction (black background response) normally evoked by transferring an animal from darkness to superior illumination in light-absorbing surroundings. Light initiates this reaction when it falls on a *B* area located in the ventral part of the retina of the vertebrate eye, the more dorsal ommatidia of the isopod and the dorso-central ommatidia of the decapod eye.

(c) A *tertiary ocular* reaction (white background response) normally evoked by transferring animals from superior illumination in light-absorbing to superior illumination in light-scattering surroundings. Light initiates the reaction when it falls on a *W* area dorsal to the *B* area of the retina of the vertebrate eye, on the latero-ventral ommatidia of the isopod eye and on the ommatidia of the ventral periphery of the decapod eye.

(ii) The primary dermal response is partly autogenous, i.e. evoked by direct action of light on the melanophores, and partly due to dermal photoreceptors which may bring into play direct nervous co-ordination of innervated chromatophores.

(iii) The secondary ocular response of all vertebrates, except possibly some teleosts, depends on *reflex* liberation of the *B* substance of the pars intermedia of the pituitary gland.

(iv) The tertiary ocular response is the latest component in the evolutionary series: and its control is different in different groups. To a greater or less extent in reptiles and in Teleostei direct nervous control has *independently* established itself as a controlling process; but in Teleostei it seems to be superimposed on a more archaic type of indirect control through reflex liberation of some hormone. Study of colour changes of Amphibia, of the eel and of elasmobranchs points to the conclusion that some part of the pituitary (possibly the *pars tuberalis*) is either the seat of the latter or controls the functional activity of another ductless gland which produces it—possibly the suprarenal.

(v) Electrical and tactile stimulation of certain receptive areas of reptiles which

have acquired direct nervous control of the pigmentary effector system can evoke a reaction like the tertiary ocular response; but electrical and tactile stimulation rarely evoke generalized chromatic response of other vertebrates.

(vi) We have insufficient information to justify any general conclusions about the way in which temperature or moisture evoke response of land vertebrates or terrestrial Crustacea; but the action of the former may be partly autogenous.

(vii) The fact that so many species of vertebrates which possess pigmentary effector organs (i.e. nearly all cold-blooded ones) display no striking colour change in response to normal changes of the environment is:

(a) partly due to specific variation of the resting level of hormone concentration in the blood;

(b) partly due to a very delicate adjustment of ocular orientation and external illumination necessary for the production of separate secondary and tertiary light reactions;

(c) partly due to specific variation of sensitivity of pigmentary effector organs to the direct action of radiation.

(viii) The control of chromatic behaviour of Crustacea is due to reflex liberation of at least two hormones, of which one is produced by a gland in the neighbourhood of the eye itself.

(ix) The sinus gland of the eye stalk of decapods produces a hormone the reflex liberation of which evokes the black background response of *Brachyura* and the white background response of decapods other than *Brachyura*.

REFERENCES

- Abramowitz, A. A. 1937^a The chromatophorotropic hormone of the Crustacea: standardization, properties, and physiology of the eye-stalk gland. *Biol. Bull. Woods Hole*, **72**, 344.
- Abramowitz, A. A. 1937^b The comparative physiology of pigmentary responses in the Crustacea. *J. Exp. Zool.* **76**, 407.
- Abramowitz, A. A. 1937^c The role of the hypophyseal melanophore hormone in the chromatic physiology of *Fundulus*. *Biol. Bull. Woods Hole*, **73**, 134.
- Allen, B. M. 1917 Effects of the extirpation of the anterior lobe of the hypophysis of *Rana pipiens*. *Biol. Bull. Woods Hole*, **32**, 117.
- Atwell, W. J. 1941 The morphology of the hypophysis cerebri of toads. *Amer. J. Anat.* **68**, 191.
- Ballowitz, E. 1893 Die Nervenendigungen der Pigmentzellen u.s.w. *Z. wiss. Zool.* **56**, 673.
- Biedermann, W. 1892 Über den Farbenwechsel der Frosche. *Pflüg. Arch. ges. Physiol.* **51**, 455.
- Brown, F. A. 1935 Control of pigment migration within the chromatophores of *Palaemonetes vulgaris*. *J. Exp. Zool.* **71**, 1.
- Butcher, E. O. 1938 The structure of the retina of *Fundulus heteroclitus* and the regions of the retina associated with different chromatophoric responses. *J. Exp. Zool.* **79**, 275.
- Carlson, S. P. 1936 Colour changes in brachyuran crustaceans, especially in *Uca pugnator*. *K. fysiogr. Sällsk. Lund Förh.* **6**, 63.
- Dawes, B. 1941 The melanin content of the skin of *Rana temporaria* under normal conditions and after prolonged light and dark adaptation. *J. Exp. Biol.* **18**, 26.
- Geiringer, M. 1935 Dioxyphenyle im Vorderlappen der Wirbeltierhypophyse. *C.R. Congr. XIIème Int. Zool. Lisbon*, p. 755.

- Giersberg, H. 1930 Über Farbwechsel der Tiere. *Jber. schles. Ges. vaterl. Kult.* 103, 25.
- Giusti, L. & Houssay, B. A. 1924 Modificaciones cutaneas y genitales en el sapo por ablación de la hipófisis o lesiones del cerebro. *Rev. Soc. argent. Biol.* 5, 155.
- Hanström, B. 1937 Die Sinusdrüse und der hormonal bedingte Farbwechsel der Crustaceen. *K. svenska VetenskAkad. Handl.* 3, 16.
- Hill, A. V., Parkinson, J. L. & Solandt, D. Y. 1935 Photo-electric records of the colour change in *Fundulus*. *J. Exp. Biol.* 12, 397.
- Hogben, Lancelot 1923 A method of hypophysectomy in adult frogs and toads. *Quart. J. Exp. Physiol.* 13, 177.
- Hogben, Lancelot 1924 The pigmentary effector system. IV. A further contribution to the role of pituitary secretion in amphibian colour response. *Brit. J. Exp. Biol.* 1, 249.
- Hogben, Lancelot 1936 The pigmentary effector system. VIII. The chromatic function in elasmobranch fishes. *Proc. Roy. Soc. B*, 120, 142.
- Hogben, Lancelot & de Beer, G. R. 1925 Studies on the pituitary. V. Localisation and phyletic distribution of active materials. *Quart. J. Exp. Physiol.* 15, 163.
- Hogben, Lancelot & Crew, F. A. E. 1923 Studies on internal secretion. II. Endocrine activity in foetal and embryonic life. *Brit. J. Exp. Biol.* 1, 1.
- Hogben, Lancelot & Gordon, C. 1930 Studies on the pituitary. VII. The separate identity of the pressor and melanophore principles. *J. Exp. Biol.* 7, 286.
- Hogben, Lancelot & Landgrebe, F. W. 1940 The pigmentary effector system. IX. Receptor fields of teleostean visual response. *Proc. Roy. Soc. B*, 128, 317.
- Hogben, Lancelot & Mirvish, L. 1928 The pigmentary effector system. V. The nervous control of excitement pallor in reptiles. *J. Exp. Biol.* 5, 295.
- Hogben, Lancelot & Slome, D. 1931 The pigmentary effector system. VI. The dual character of endocrine coordination in amphibian colour change. *Proc. Roy. Soc. B*, 108, 10.
- Hogben, Lancelot & Slome, D. 1936 The pigmentary effector system. VII. The dual receptive mechanism of amphibian background response. *Proc. Roy. Soc. B*, 120, 158.
- Hogben, Lancelot & Winton, F. R. 1922a The pigmentary effector system. I. *Proc. Roy. Soc. B*, 93, 318.
- Hogben, Lancelot & Winton, F. R. 1922b The pigmentary effector system. II. *Proc. Roy. Soc. B*, 94, 151.
- Hogben, Lancelot & Winton, F. R. 1922c Studies on the pituitary. I. The melanophore stimulant in posterior lobe extracts. *Biochem. J.* 16, 619.
- Hogben, Lancelot & Winton, F. R. 1923 The pigmentary effector system. III. *Proc. Roy. Soc. B*, 95, 15.
- Houssay, B. A., Biasotti, A. & Sammartino, R. 1935 Modificaciones producidas por las lesiones infundibulo-luberianas en el sapo. *Rev. Soc. argent. Biol.* 11, 318.
- Houssay, B. A. & Ungar, I. 1925 Factores que regulan la pigmentación de la Rana *Leptodactylus ocellatus*. *Rev. Soc. argent. Biol.* 6, 5.
- Keeble, F. & Gamble, F. W. 1904 The colour physiology of higher Crustacea. II. *Phil. Trans.* 196, 295.
- Kleinholz, L. H. 1935 The melanophore dispersing principle in the hypophysis of *Fundulus heteroclitus*. *Biol. Bull. Woods Hole*, 69, 379.
- Kleinholz, L. H. 1937 Studies in the pigmentary system of Crustacea. II. Diurnal movements of the retinal pigments of Bermudan decapods. *Biol. Bull. Woods Hole*, 72, 176.
- Kleinholz, L. H. 1938a Studies in the pigmentary system of Crustacea. IV. The unitary versus the multiple hormone hypothesis of control. *Biol. Bull. Woods Hole*, 75, 510.
- Kleinholz, L. H. 1938b Studies in reptilian colour changes. II. The pituitary and adrenal glands in the regulation of melanophores of *Anolis carolinensis*. *J. Exp. Biol.* 15, 474.
- Kleinholz, L. H. 1938c Studies in reptilian colour changes. III. Control of the light phase and behaviour of isolated skin. *J. Exp. Biol.* 15, 492.
- Koller, G. 1927 Über chromatophorensystem, Farbensinn u. Farbwechsel bei *Crangon vulgaris*. *Z. vergl. Physiol.* 5, 191.
- Koller, G. 1929 Versuche über die inkretorischen Vorgänge beim Garneelenfarbwechsel. *Z. vergl. Physiol.* 8, 601.
- Krogh, G. 1922 *The anatomy and physiology of capillaries*. Yale Univ. Press.

- Landgrebe, F. W. & Waring, H. 1941 Intermediate lobe pituitary hormone. *Quart. J. Exp. Physiol.* **31**, 31.
- Laurens, H. 1915 The reactions of the melanophores of *Amblystoma* larvae. *J. Exp. Zool.* **18**, 577.
- Laurens, H. 1916 The reactions of the melanophores of *Amblystoma* larvae. The supposed influence of the pineal organ. *J. Exp. Zool.* **20**, 237.
- Laurens, H. 1917 The reactions of the melanophores of *Amblystoma tigrinum* larvae to light and darkness. *J. Exp. Zool.* **23**, 195.
- Lundstrom, H. M. & Bard, P. 1932 Hypophyseal control of cutaneous pigment in elasmobranch fish. *Biol. Bull. Woods Hole*, **62**, 1.
- Martini, E. & Achundow, I. 1929 Versuche über Farbenpassung bei Culiciden. *Zool. Anz.* **81**, 25.
- Matthews, S. A. 1931 Observations on pigment migration within the fish melanophore. *J. Exp. Zool.* **58**, 471.
- Matthews, S. A. 1933 Colour changes in *Fundulus* after hypophysectomy. *Biol. Bull. Woods Hole*, **64**, 315.
- Neill, R. M. 1940 The existence of two types of chromatic behaviour in teleostean fishes. *J. Exp. Biol.* **17**, 74.
- Neill, R. M. 1942 The eye of *Salmo salar* (in the Press).
- Osborn, C. M. 1938 The role of the melanophore dispersing hormone of the pituitary in the colour changes of the catfish. *J. Exp. Zool.* **79**, 309.
- Parker, G. H. 1903 The skin and eyes as receptor organs in the reactions of frogs to light. *Amer. J. Physiol.* **10**, 28.
- Parker, G. H. 1906 The influence of light and heat on the movement of the melanophore pigment, especially in lizards. *J. Exp. Zool.* **3**, 401.
- Parker, G. H. 1938 The colour change in lizards, particularly in *Phrynosoma*. *J. Exp. Biol.* **15**, 48.
- Parker, G. H. 1941 Melanophore bands and areas due to nerve cutting, etc. *J. Gen. Physiol.* **24**, 483.
- Parker, G. H. & Rosenbluth, A. 1941 Electrical stimulation of the concentrating (Adrenergic) and the dispersing (cholinergic) nerve fibres of the melanophores of the catfish. *Proc. Nat. Acad. Sci., Wash.*, **27**, 198.
- Perkins, E. B. 1928 Colour changes in crustaceans especially in *Palaemonetes*. *J. Exp. Zool.* **50**, 71.
- Perkins, E. B. & Snook, T. 1932 The movement of pigment within chromatophores of *Palaemonetes*. *J. Exp. Zool.* **61**, 115.
- Redfield, A. C. 1918 Physiology of the melanophores of *Phrynosoma*. *J. Exp. Zool.* **26**, 275.
- Smith, H. G. 1920 The pigmentary and endocrine disturbances induced in the anuran tadpole by the early ablation of the pars buccalis of the hypophysis. *Amer. Anat. Mem.* no. **11**.
- Smith, H. G. 1938 The receptive mechanism of the background response in chromatic behaviour of Crustacea. *Proc. Roy. Soc. B*, **125**, 250.
- Ståhl, F. 1938 Über das Vorkommen von inkretorischen Organen u. Farbwechselshormone im Kopf einiger Crustaceen. *K. fysiogr. Sällsk. Lund Förh.* N.F. **49**. Preliminary report on the colour changes and the incretory organs in the heads of some crustaceans. *Ark. Zool.* **30 B**, no. 8.
- Steggerda, F. R. & Soderwall, A. L. 1939 Relationship of the pars tuberalis to melanophore response in Amphibia. *J. Cell. Comp. Physiol.* **13**, 31.
- Stoppani, A. O. M. 1941 Estudios fisiologicos y farmacologicos sobre los melanoforos de los Batracios. *Trab. Inst. de Fisiol. de la Facult. de Ciencias Medic.*
- Sumner, F. B. 1933 The differing effects of differing parts of the visual field upon the chromatophore responses of fishes. *Biol. Bull. Woods Hole*, **65**, 266.
- Swingle, W. W. 1921 The relation of the pars intermedia of the hypophysis to pigmentation changes in anuran larvae. *J. Exp. Zool.* **34**, 119.
- Waring, H. 1938 Chromatic behaviour of elasmobranchs. *Proc. Roy. Soc. B*, **125**, 264.
- Waring, H. 1940 The chromatic behaviour of the eel. *Proc. Roy. Soc. B*, **128**, 343.

- Waring, H. & Landgrebe, F. W. 1941 On chromatic effector speed in *Xenopus* and *Anguilla* and the level of melanophore expanding hormone in eel blood. *J. Exp. Biol.* **18**, 80.
- Waring, H., Landgrebe, F. W. & Bruce, J. R. 1942 Chromatic behaviour of *Scyllium canicula*. *J. Exp. Biol.* **18**, 306.
- Wells, G. P. 1932 Colour response in a leech. *Nature, Lond.*, **129**, 686.
- Wykes, U. 1936 Pigmentary coordination in elasmobranchs. *J. Exp. Biol.* **13**, 460.
- Young, J. Z. 1935*a* The photoreceptors of the lamprey. I. Light-sensitive fibres in the lateral line nerve. *J. Exp. Biol.* **12**, 229.
- Young, J. Z. 1935*b* The photoreceptors of the lamprey. II. The functions of the pineal complex. *J. Exp. Biol.* **12**, 254.
- Zoond, A. & Bokenham, N. A. H. 1935 Studies in reptilian colour response. II. The role of retinal and dermal photoreceptors in the pigmentary activity of the chameleon. *J. Exp. Biol.* **12**, 39.
- Zoond, A. & Eyre, J. 1934 Studies in reptilian colour response. I. The binomics and physiology of the pigmentary activity of the chameleon. *Phil. Trans. B*, **223**, 27.

BAKERIAN LECTURE*

Amino-acid analysis and the structure of proteins

BY A. C. CHIBNALL, F.R.S.

(Delivered 16 July 1942—Received 27 August 1942)

INTRODUCTION

In recent years the X-ray crystallographers have made remarkable advances in the interpretation of protein structure, and it is becoming more and more evident that a stage has been reached when their views need to be reconciled with data obtained from accurate amino-acid analysis of the proteins concerned. In all too many cases these data are, unfortunately, not yet available, and the reason why the analyst cannot supply them at short notice is due not so much to the complexity of the problem—which he has never sought to minimize—but to the fact that many of the more important methods of analysis in current use are an inheritance from an earlier period when such accuracy as is now demanded would have been considered almost impossible of achievement.

From about 1840 until 1900, following the lead given by Liebig and later by Ritthausen, the attention of protein chemists was centred chiefly on the preparation and characterization of various animal and seed proteins; as substances of physiological interest their enzymic digestion products were studied in elaborate detail by Kühne, but little attention was paid to the ultimate decomposition

* The full data for the free amino-groups in edestin given in table 9 of the present paper were not available at the time the abstract of the lecture went to press (*Nature, Lond.*, 1942, **150**, 127). This accounts for the slightly different views on the structure of edestin now presented.

products, the amino-acids, in spite of the fact that Ritthausen as early as 1872 had pointed out that the proportions in which these occur might be characteristic of the protein concerned. The enunciation by Hofmeister and Fischer of the peptide hypothesis in 1901 emphasized for the first time the fundamental importance of the amino-acids, and a most fruitful period followed in which attention became almost exclusively focused on these products. Under the inspiring leadership of Fischer himself great improvements were effected in the separation and identification of the amino-acids, so that by about 1915 reasonably good analyses were available for most of the better-known proteins. Though far from complete, the analytical data showed quite clearly that proteins could differ widely in composition, and in many cases it was possible to correlate composition with nutritive value. Such an aim was, indeed, the incentive behind much of the work of this period.

During the next 15 years interest remained centred around the amino-acids, and further improvements were effected in the separation of the bases, the dicarboxylic acids and to a lesser extent the monoamino-acids. Progress, however, was slow, for but few workers were attracted to the field; although in certain cases some 90 % of the hydrolysis products had been recognized, analysts of experience were still unwilling to admit that their data, even for the bases and dicarboxylic acids, were quantitative. This was because most of the gravimetric estimations made use of selective precipitants and were thus at the mercy of solubility factors whose magnitude could vary from protein to protein according to the amino-acid mixture present in the hydrolysate. Indirect and specific methods of estimation, such as those for tyrosine, tryptophan and cystine, were not open to these objections, but were susceptible to extraneous errors.

Around 1930, chiefly from lack of the necessary incentive, interest in amino-acid analysis had waned and the study of the proteins themselves came once more to the fore. Measurements of osmotic pressure and of ultracentrifugal sedimentation were showing that the proteins had molecular weights running into many tens and in some cases many hundreds of thousands, so that the gap between the intact molecule and its decomposition products seemed to be widening. The long peptide chain of Hofmeister and Fischer became if anything more of an abstract conception than ever, in spite of the fact that Astbury's X-ray studies of wool keratin were providing convincing proof of its reality. Something was missing; something that would enable us to perceive, however dimly, the way in which nature builds when she makes a mountain of a molehill, for if these gigantic molecules were indeed chemical entities their structure must ultimately be interpreted in terms of their building stones, the constituent amino-acids. Fortunately for the chemist, Astbury (1934), in a paper that displays remarkable intuition, provided the missing link when he introduced the concept of the hypothetical average amino-acid residue weight, for this factor enables us to interpret our analytical data in terms of the total number of residues in the molecule.

Astbury was discussing the structure of gelatin, and it was necessary for him to

know the average weight of protein associated with the average residue dimensions given by X-rays. He accordingly calculated the molecular weight of the hypothetical average amino-acid residue (referred to henceforth as the mean residue weight) by deducing first of all, from known analytical data, the number of gram residues of the average amino-acid. His tabulated values for the number of gram residues of the various amino-acids in gelatin revealed the striking fact, which he had undoubtedly anticipated from a study of the X-ray photographs of the protein, that the residues of glycine and hydroxyproline account, respectively, for about one-third and one-ninth of the total number of residues, that is to say, every third residue could be a glycine residue and every ninth a hydroxyproline residue.

This latter suggestion, implying that the two residues occur at regular intervals along the peptide chain, was extended a little later by Bergmann & Niemann (1936, 1937, 1938), who assumed that the periodicity noted for these two residues in gelatin was true also for every species of residue in all proteins, in other words, that the occurrence of residues in the peptide chain at regular intervals was a fundamental law of protein structure. Applying this generalization to their own experimental data for gelatin, fibrin, cattle globin and silk fibroin, and to the data of other workers for egg albumin, they drew the further important conclusion that the numbers denoting frequency of recurrence of the residues, and also the number of total residues in the molecule, are not random numbers but can all be expressed in powers of two and three. The number of residues in the proteins they had examined fell into a series of multiples of 288; for example, egg albumin had 288 ($2^5 \times 3^2$), cattle globin and fibrin 576 ($2^6 \times 3^2$), and silk fibroin 2592 ($2^5 \times 3^4$) residues respectively. By multiplying these numbers by the respective average residue weights they obtained values of 35,700, 66,520, 69,300, and 217,000 for the corresponding molecular weights. All these findings pointed in a most striking way to underlying regularities in the building up of protein molecules and were in harmony with the recent views of Svedberg (1938) based on observations of particle size. To those interested in the subject it seemed like the dawn of a new era, and as Bergmann (1938) himself so appositely remarked: 'Everyone who is familiar with the history of protein chemistry may feel somewhat amazed on being confronted with a simple stoichiometry of the protein molecule.' Some of us, nevertheless, were not prepared to give this attractive generalization our immediate and unqualified support, for we questioned the reliability and completeness of the amino-acid analyses on which it was based:

Oh, the little more, and how much it is!
And the little less, and what worlds away!

Neuberger (1939) suggested that the average error in the determination of individual amino-acids in a protein might be fairly taken as 5-6 % and, assuming the possibility of an error of 6 %, he calculated that there was a very high degree of probability (approx. 80 %) that a purely random distribution of amino-acids

should give values in apparent accordance with the formula of the frequency hypothesis. A similar view was voiced by Pirie (1939) who pointed out that in a series of numbers expressible as $2^n \times 3^m$ (where n and m are positive whole numbers or zero) the only long gaps are 19-23 and 37-47. The other intervals are small compared with the probable error in most amino-acid estimations on protein hydrolysates, so that it is the absence of figures such as 21 and 42 in the list of values for the number of occurrences of an individual amino-acid in the protein molecule that tests the hypothesis.

Having been interested for some years in the isolation of asparagine and glutamine from proteins my own attitude towards the hypothesis was quite naturally influenced by the fact that Bergmann & Niemann had assigned to aspartic and glutamic acids frequencies that were based on the amounts isolated after hydrolysis; this seemed to me an unjustifiable procedure, for in each case these amounts represent two residues, those of the amide and of the free amino-acid respectively. Pirie's (1939) criticism of this was to the point: 'In general the sum of two numbers expressible as $2^n \times 3^m$ is not itself expressible in this way; it is therefore curious that the glutamic acid content of several proteins has satisfied the numerical generalizations if it arises from two independent amino-acid frequencies.' In their dicarboxylic acid analyses Bergmann & Niemann (also Calvery with egg albumin) had employed the method of Foreman, which was one that my colleagues and I had had under investigation in my laboratory for some time; we had in fact already introduced important modifications which had enabled us to obtain values for the aspartic and glutamic acid contents of haemoglobin (horse and cattle) and of egg albumin that were much in excess of those used by Bergmann & Niemann in their frequency computations; we had evidence, moreover, that our values were not yet maximal. Although these findings had no direct bearing on the general validity of the hypothesis, for this could not apply to them without reference to the amide-ammonia, they did reinforce my original surmise that our methods of amino-acid analysis needed a searching reinvestigation before they could be used with confidence to confirm or reject such very important conclusions as those of Bergmann & Niemann. Gravimetric methods of estimation needed specially close attention, for they were open to the objection that the various precipitants used were giving low values on account of adventitious solubility factors. Bergmann himself was clearly worried by the same disturbing thoughts, for he initiated in 1938 new methods, which he is still developing, to overcome such possible solubility errors.

Our approach to the problem has been on lines which have necessitated no fundamental departure from existing methods, and the scheme of analysis has taken us about five years to bring to fruition. Our aim has been to isolate by step-wise removal without overall loss, as pure derivatives and in quantitative yield, aspartic acid, glutamic acid, arginine, histidine and lysine, for all of these contain strongly ionizable groups, and as residues their number and arrangement in the protein molecule must to a certain extent control its specific polypeptide pattern and be responsible for its behaviour in solution. The cardinal point in our analytical

procedure is that no reagent may be introduced into the hydrolysate unless it can be removed at a later stage without loss. Full details of our methods will be published in the near future, and all that I need add here is that the solubility factors inherent in the Foreman treatment for the dicarboxylic acids, and in the phosphotungstic acid precipitation of the bases, are overcome by repeating these operations at a later stage, when about four-fifths of the monoamino-acids and proline have been removed from the collected mother liquors. The whole procedure, which requires some months to carry through to completion, is admittedly very laborious and demands analytical skill of a high order; yet the results for egg albumin, edestin and β -lactoglobulin have been worth while, for overall losses of nitrogen have been reduced to 3, 2.5 and 1.5 % respectively, and of this the major part can be shown from flow sheets to fall on the monoamino-acids. I am confident that our estimates for the aspartic acid, glutamic acid, arginine, histidine and lysine contents of these three proteins are within 1 or 2 % of the true values, and that these proteins do not contain any hydroxyglutamic acid, hydroxylysine or unrecognized diamino-acids. The values for tyrosine, tryptophan, cystine and methionine quoted in the discussions which follow are from, or are based on, the recent researches of Kassell & Brand, Baernstein and Lugg, and are believed to be accurate to within 2 or 3 %. It is doubtful if the value for cysteine given in table 3 can be considered as falling within these limits.

THE MEAN RESIDUE WEIGHT OF PROTEINS

Before making use of all these new analytical data in an attempt to interpret protein structure I must first deduce a value for the mean residue weight of the protein concerned. Astbury (1934) estimated this in the following way. From known analytical data he calculated for each amino-acid the number of gram residues in 100 g. of protein, a corresponding value for the remaining acids, assumed to be monoamino-acids, being obtained indirectly from a consideration of the total nitrogen content of the protein and of the acids already accounted for. By summation he was thus able to show that 100 g. of gelatin contained approximately 1.044 g. residues of the average amino-acid residue, so that the molecular weight of the latter was $100/1.044 = 96$. Bergmann and his colleagues have preferred to give greater weight to the known analytical data, and in their computations for cattle globin (115.5), egg albumin (124), blood fibrin (120.3), silk fibroin (84) and gelatin (93.7) they have taken into account the fact that methods of analysis for certain of the amino-acids, especially the simpler monoamino-acids, are far from quantitative (Bergmann & Niemann 1937, 1938; Bergmann & Stein 1939). The procedure that I propose to use is in essence the same as that of Astbury, but the amount of free and potential (i.e. peptide) α -amino (including imino)-N, which is a measure of the total number of residues, is computed indirectly from data giving the amount of nitrogen present in other forms.

If the percentage of total nitrogen in the protein is N and the percentage of this

nitrogen in the form of free and potential α -amino (including imino)-N is M , then 100 g. of protein will contain $N \times M/100$ g. of free and potential α -amino (including imino)-N. Designating the number of g. equivalents of free and potential α -amino (including imino)-N per 100 g. as G , we have

$$G = \frac{N \times M}{14 \times 100}. \quad (1)$$

Let R be the mean residue weight, then

$$R = \frac{100}{G} = \frac{100 \times 100 \times 14}{N \times M}. \quad (2)$$

If S , A , H , L , L' and T are the percentages of total protein-N present as amide-N, arginine-N, histidine-N, lysine-N, hydroxylysine-N and tryptophan-N respectively, then

$$M = 100 - S - \frac{3A}{4} - \frac{2H}{3} - \frac{L}{2} - \frac{L'}{2} - \frac{T}{2}. \quad (3)$$

From equations (2) and (3) therefore the mean residue weight can be calculated in terms of the total nitrogen and of that of the amino-acids which contain non- α -amino (including imino)-N. This evaluation of R will be completely valid if we assume that the protein molecule is a polypeptide chain in which all the peptide linkages are concerned with α -amino (including imino)-groups, or a system of such chains in which each cross-linkage is formed by a condensation involving the elimination of a molecule of water. Moreover, the calculation will remain significant provided that the amount of any other cross-linkage (e.g. disulphide) or of undetermined amino-acid residues with nitrogen in excess of that required for the peptide linkage (e.g. citrulline) is small. My own experience, which I believe is in keeping with that of other protein analysts, suggests that unrecognized linkages or residues of this type are not present in any large amount, though further research may perhaps show that this statement is not true for unique proteins from organs with specialized functions. It should be noted that small errors in the determination of the factors concerned in equation (3) will not significantly affect the evaluation of R ; it is essential nevertheless that N be accurately determined.

By a similar procedure it is possible to evaluate W , the weight of hydrolysis products given by 100 g. of protein, for each peptide and amide linkage will bind one molecule of water, so that

$$W = 100 + \frac{18N(P+S)}{14 \times 100}, \quad (4)$$

where P is the percentage total protein-N present in peptide linkages. Now the total N present in free terminal amino (including imino)-groups will be one equivalent for each peptide chain in the system. A determination of free amino-N on the intact protein by the method of Van Slyke (assuming no secondary reactions to

occur, a point that will be discussed later) will give the free terminal amino-N, half the lysine-N and half the hydroxylysine-N. Hence, if V is the percentage of total protein-N given by a Van Slyke determination on the intact protein, then, neglecting possible free terminal imino-groups

$$P = 100 - S - \frac{3A}{4} - \frac{2H}{3} - \frac{T}{2} - V. \quad (5)$$

In addition to this we can calculate Z , the weight of amino-acids given on hydrolysis of 100 g. of protein, for the molecular weights of asparagine and glutamine differ only by one from those of the respective acids,

$$Z = 100 + \frac{18N \times P}{14 \times 100}. \quad (6)$$

If the number of peptide linkages is large compared with the number of free terminal amino (including imino)-groups, so that as a first approximation one can assume that P is equal to M , then, from equations (2) and (6),

$$Z = \frac{100(R + 18)}{R}. \quad (7)$$

Computed values for R , W , and Z for a number of proteins are given in table 1. It will be noted that the value of R for β -lactoglobulin is in excellent agreement with that of Hotchkiss (1939), who determined the number of peptide linkages split by enzymic hydrolysis. As I had anticipated, my values for egg albumin, cattle globin and silk fibroin are significantly lower than those computed by Bergmann and his colleagues. Turning to the computed values for Z it is interesting to note that proteins of such very different composition as edestin, gliadin and β -lactoglobulin give very nearly the same weight of amino-acids on hydrolysis; this is, of course, because they have approximately the same value for R . Z must decrease as R increases, but as Dr Astbury has pointed out to me, the changes in Z will be only about one-seventh of the corresponding change in R , for, differentiating equation (7),

$$\delta Z = \frac{-1800}{R^2} \delta R,$$

and if R be of the order of 115

$$\delta Z = \frac{-1800}{13,225} \delta R = -\delta R/7 \text{ (approx.)}.$$

This means that in any experiments to determine R by measuring Z (or W) directly, as we have been able to do incidentally during our analyses of edestin, egg albumin and zein, Z (or W) would have to be determined very accurately.

TABLE 1. THE MEAN RESIDUE WEIGHT OF VARIOUS PROTEINS

protein	as percentages of total protein-N										calculated weight of amino-acids given on hydrolysis of 100 g. protein Z ⁽¹²⁾	
	N percentage total protein-N ⁽¹¹⁾	A arginine-N	H histidine-N	L lysine-N	L' hydroxy-lysine-N ⁽¹⁰⁾	T tryptophan-N	S amide-N ⁽¹¹⁾	A free-amino-N	mean residue weight	weight of hydrolysis product per 100 g. protein		
										calculated		found
									R	W ⁽¹²⁾		
									calculated	literature	calculated	
edestin	18.65	28.85	3.50	2.44	—	1.09	9.49	1.52	115.7	—	117.8	117.2
egg albumin	15.76	11.46	2.50	6.15	0.1	1.15	6.4	4.5	111.4	124 ⁽⁵⁾	117.2	117.3
β -lactoglobulin	15.58	5.95	2.69	12.0	—	1.71 ⁽⁴⁾	6.85	7.9	112.4	113 ⁽¹⁴⁾	117.0	—
myosin	16.8	13.4 ⁽⁶⁾	2.8 ⁽⁶⁾	11.76 ⁽⁶⁾	—	0.66 ⁽⁷⁾	7.1	—	111.3	116 ⁽¹⁷⁾	117.8	—
										96 ⁽¹¹⁾		
gelatin	18.27	15.3 ⁽⁶⁾	1.33 ⁽⁸⁾	6.2 ⁽⁶⁾	0.9	0.0	0.5	3.1 ⁽¹²⁾	91.6	93.7 ⁽¹²⁾	119.8	—
casein	15.65	8.0 ⁽¹²⁾	3.2 ⁽¹²⁾	7.7 ⁽¹²⁾	0.33	1.17	9.05	5.5 ⁽¹²⁾	114.6	—	117.3	—
gliadin	17.66	5.0 ⁽¹¹⁾	2.77 ⁽¹²⁾	0.7 ⁽¹³⁾	0.12	0.5	25.35	1.0 ⁽¹²⁾	115.8	—	121.1	—
zein	16.2	3.24 ⁽⁹⁾	1.97 ⁽¹²⁾	0.0	0.33	0.14	18.3	0.0 ⁽¹²⁾	111.2	—	120.0	119.7
wool keratin	16.57	20.2 ⁽¹⁰⁾	1.23 ⁽¹⁰⁾	3.12 ⁽¹⁰⁾	0.11	0.62 ⁽¹⁶⁾	7.2	—	113.0	118 ⁽¹⁷⁾	117.5	—
insulin	15.64	6.37	19.1	1.57	—	0.0	8.9	4.6	123.8	—	113.8	—
silk fibroin	19.0	1.30 ⁽¹³⁾	0.1 ⁽¹³⁾	0.25 ⁽¹³⁾	—	0.0	2.0	—	76.1	84 ⁽¹⁸⁾	124.1	—
horse globin	17.12	6.63 ⁽¹²⁾	12.56 ⁽¹²⁾	9.43 ⁽¹²⁾	0.0	1.70	4.58	6.56	106.9	—	117.4	—
cattle globin	17.0	5.9 ⁽⁶⁾	11.8 ⁽⁶⁾	9.0 ⁽⁶⁾	—	1.86 ⁽¹¹⁾	4.27	6.0 ⁽¹²⁾	105.5	115.5 ⁽¹³⁾	117.7	—

(1) All values quoted on an ash-free basis.

(2) All values except that for silk fibroin were obtained by the method of Gordon, Martin & Syngé (1941).

(3) Macpherson (unpublished).

(4) Brand & Kassell (1942).

(5) Bergmann & Niemann (1937).

(6) Sharp (1939).

(7) Bailey (1937).

(8) Vickery (1940).

(9) Dakin (1920).

(10) Van Slyke, Hiller & MacFadyen (1941).

(11) Astbury (1934).

(12) Bergmann & Stein (1939).

(13) Vickery & White (1933).

(14) Hotchkiss (1939), on basis of $N = 15.55$

instead of 15.3 %.

(15) Osborne, Van Slyke, Leavenworth & Vino-

grad (1915).

(16) Block (1939).

(17) Astbury (1942).

(18) Bergmann & Niemann (1938).

(19) Vickery & Block (1931).

(20) Van Slyke & Birchard (1913).

(21) Roche & Jean (1934).

(22) Vickery & Leavenworth (1928).

(23) Where no values for V are quoted W and Z have been calculated on the assumption that P is equal to M (see text).

THE STRUCTURE OF EDESTIN

Having computed the mean residue weight I am now in a position to interpret our new analytical data for edestin (table 2). The mean residue weight is 115.7 (table 1), so that 1 g. of protein contains $1/115.7$ or 864×10^{-5} g.mol. of the mean residue. Column 2 shows for each amino-acid the number of g.mol. $\times 10^{-5}$ present in 1 g. protein calculated from the analytical values in column 1. The frequency

TABLE 2. EDESTIN

- (1) Percentage of amino-acid given on hydrolysis. Values for tyrosine, tryptophan, cystine and methionine are from Lugg (1938 *a, b*); the remainder are from unpublished analyses.
- (2) Number of g.mol. $\times 10^{-5}$ per g. protein calculated from the data in (1).
- (3) Number of g.mol. $\times 10^{-5}$ per g. protein calculated from the data in (4).
- (4) Frequencies calculated from the number of g.mol. $\times 10^{-5}$ per g. protein (864) of the mean residue and the data in (2), adjusted to the nearest integral factor of 864.
- (5) Number of residues in a unit containing 432 residues.

amino-acid:	(1)	(2)	(3)	(4)	(5)
arginine	16.71	96.0	96.0	9 (3 ²)	48 (2 ⁴ × 3)
histidine	2.41	15.6	16.0	54 (2 × 3 ³)	8 (2 ³)
lysine	2.37	16.2	16.0	54 (2 × 3 ³)	8 (2 ³)
tyrosine	4.34	24.0	24.0	36 (2 ³ × 3 ²)	12 (2 ³ × 3)
tryptophan	1.50	7.35	8.0	108 (2 ³ × 3 ³)	4 (2 ²)
methionine	2.35	15.8	16.0*	54 (2 × 3 ³)	8 (2 ³)
cystine	1.44	6.0	6.0	144 (2 ⁴ × 3 ²)	3 (3)
			(L.C.M.)	432 (2 ⁴ × 3 ³)	
glutamic acid	20.70	141.0	144.0	6 (2 × 3)	72 (2 ³ × 3 ²)
aspartic acid	12.0	90.0	90.0	9.82	45
ammonia	2.15	126.5	126.0	6.86	63
as residues before hydrolysis:					
glutamic acid			72.0	12 (2 ² × 3)	36 (2 ² × 3 ²)
glutamine			72.0	12 (2 ² × 3)	36 (2 ² × 3 ²)
aspartic acid			36.0	24 (2 ³ × 3)	18 (2 × 3 ²)
asparagine			54.0	16 (2 ⁴)	27 (3 ³)

* Osborne (1902*b*) found 0.88 % of protein-S in edestin: the values for methionine and cystine in column 3 require 0.886 % of protein-S.

of each amino-acid is determined by dividing the total number of g.mol. of the mean residue per g. protein, namely, 864, by the appropriate figure in column 2; in column 4 these frequencies are shown after adjustment to the nearest integral factor of 864. The theoretical figures for the g.mol. $\times 10^{-5}$ per g. protein of the individual amino-acids corresponding to these smoothed frequencies appear in column 3. Agreement between columns 2 and 3 is good in all cases except for tryptophan, where the observed value is somewhat low. The frequency numbers given in column 4 are all factorizable by 2 and/or 3, and the lowest common multiple is $2^4 \times 3^3$ or 432. So far, therefore, our analysis is well in keeping with the postulate of Bergmann & Niemann, and on the assumption that the smallest unit

of edestin contains 432 residues I have calculated the number of individual residues (column 5).

Next, let me consider the position of the dicarboxylic acids and ammonia, for I have here an example of the fallacy of treating these three hydrolysis products as simple residues. Calculation shows that in a unit of 432 residues there must be present 63 residues of amides and 54 residues of free dicarboxylic acids. The presence of asparagine residues in edestin has been experimentally verified, so that I am concerned here with residues of asparagine, glutamine, glutamic acid and perhaps also aspartic acid. If, in keeping with what has been found for the other residues, the numbers of each of these, per unit of 432 residues, are severally to be factorizable by 2 and/or 3, then calculation shows that there is indeed a unique solution to the distribution of the amide-ammonia which is that recorded in column 5. Granted the validity of the Bergmann-Niemann hypothesis therefore I have been able to estimate for the first time the amount of asparagine and glutamine residues in a protein. In this connexion the findings of Damodaran (1932) are significant. From an enzymic digest of edestin which still contained much peptide material he isolated asparagine equivalent to 13 % of the total amide-N or 9.3 asparagine residues. He also found that at the end of the digestion period the digest contained glutamine amide-N equivalent to 15.7 % and free ammonia-N equivalent to 21.3 % of the total amide-N respectively. Now the researches of Melville (1935) and of Ayre (1935) show conclusively that this ammonia arises through the incipient decomposition in aqueous solution at the temperature of the digest (37°) of glutamine and labile glutamine peptides, asparagine and its peptides being stable, so that we are justified in assuming that the glutamine amide-N produced in the digest was $15.7 + 21.3 = 37$ % of the total amide-N, which is equivalent to 26.4 glutamine residues. As the sole object of Damodaran's research was to obtain evidence for the presence of asparagine in the digest, these asparagine and glutamine residue numbers must be far from quantitative; the relative values are, nevertheless, in keeping with those I have deduced by applying the Bergmann-Niemann principle to our own dicarboxylic acid and amide-ammonia data.

There are thus reasonable grounds for assuming that our analysis of edestin, which admittedly covers only about one-half the total number of residues, provides good evidence for the Bergmann-Niemann hypothesis, and it would appear that edestin contains $2^4 \times 3^3 = 432$ amino-acid residues or a multiple thereof, so that the molecular weight of the smallest possible unit is $432 \times 115.7 = 50,000$. An association of six such units would give a molecular weight of 300,000 which is in good agreement with the value 309,000 computed from sedimentation velocity and diffusion measurements (Svedberg & Pedersen 1940). It should be noted that Burk & Greenberg (1930) observed an apparent molecular weight in 6.66*M* urea solution of 49,000, which suggests that on denaturation the association of the six component units is broken. All this evidence dovetails in very neatly with the structure I have deduced, but other data to be discussed later cannot be so

readily reconciled. This is due in large part to the fact that the insolubility of edestin at low temperature and low ionic strength makes it difficult to determine with certainty whether it is a homogeneous protein in the Tiselius apparatus and to place much reliance on the published titration data.

STRUCTURE OF β -LACTOGLOBULIN

The solubility of this protein complies with the requirements of the phase rule; this is the most sensitive test for homogeneity, and the protein has also been shown to be homogeneous in both the ultracentrifuge and the Tiselius apparatus. Our analytical data, together with some recent determinations of Brand & Kassell (1942), are given in table 3. The mean residue weight is 112.4 (table 1), so that 1 g. of protein contains 890×10^{-5} g.mol. of the mean residue; from this the frequencies

TABLE 3. β -LACTOGLOBULIN

- (1) Percentage of amino-acid given on hydrolysis. The values for tyrosine, tryptophan, cystine, cysteine and methionine are from Brand & Kassell (1942); the remainder are from unpublished analyses.
- (2) Number of g.mol. $\times 10^{-5}$ per g. protein calculated from the data in (1).
- (3) Frequencies calculated from the number of g.mol. $\times 10^{-5}$ per g. protein (890) of the mean residue, and the data in (2).
- (4) Minimal molecular weight calculated from the data in (1).
- (5) Assumed number of residues.
- (6) Calculated molecular weight of protein.

	(1)	(2)	(3)	(4)	(5)	(6)
arginine	2.89	16.6	53.6	6,015	7	42,105
histidine	1.54	9.9	89.8	10,080	4	40,290
lysine	9.75	66.7	13.3	1,498	28	41,970
tyrosine	3.78	20.9	42.7	4,790	9	43,110
tryptophan	1.94	9.5	93.7	10,520	4	42,090
methionine	3.22	21.6	41.3	4,633	9	41,700
cystine	2.29	9.5	93.5	10,500	4	41,980
cysteine	1.10	9.2	97.3	10,920	4	43,670
glutamic acid	21.51	14.6	60.6	684	62	42,400
aspartic acid	9.88	74.3	11.6	1,346	31	41,730
amide ammonia	1.29	75.9	11.7	1,318	32	42,160

given in column 3 have been calculated. There is no evidence here for the Bergmann-Niemann hypothesis, and I have accordingly calculated the minimum molecular weight which will satisfy all the analytical data. The closest approximation is 42,000, the same value as Brand & Kassell computed from their limited data and in agreement with that found by Pedersen (1936) from sedimentation velocity and diffusion measurements. The number of residues (column 5) for lysine and arginine are not factorizable by 2 and/or 3, nor is the computed total number of residues, 373. Paradoxically therefore β -lactoglobulin, which has the highest credentials for homogeneity, as emphatically refutes the Bergmann-Niemann hypothesis as edestin, about whose homogeneity nothing is known for certain,

would appear to support it. What is the solution to the enigma? Fortunately, there is further evidence to call upon in the case of β -lactoglobulin, as a careful study of the acid-base binding capacity has been made recently by Cannan, Palmer & Kibrick (1942).

Adopting a molecular weight of 42,000 and recalculating the data of these workers (based on 40,000) accordingly, I arrive at the estimates for titratable groups given in table 4, where they are set out alongside the corresponding estimates from our amino-acid analysis.

TABLE 4. ESTIMATES OF IONIZABLE GROUPS IN β -LACTOGLOBULIN

(Molecular weight taken as 42,000)

from data of Cannan <i>et al.</i> (1942)		from amino-acid analysis of protein		
ionizable groups	equiv. per mol.	amino-acid	ionizable groups	equiv. per mol.
guanidine	7	arginine	guanidine	7
imidazole	> 6	histidine	imidazole	4
amino	35-37	lysine	ϵ -amino	28
amino-N by method of Van Slyke	37	aspartic acid	31	
		glutamic acid	62	
			93	
		ammonia	32	
			61	
carboxyl	62-63		carboxyl	61

If we accept the prevailing view that the protein molecule is a simple polypeptide chain with terminal free carboxyl- and amino-groups, then the number of free carboxyl-groups that we can compute from analysis is $61 + 1 = 62$, in excellent agreement with the value found by titration. The number of free amino-groups, however, will be only $28 + 1 = 29$, some 7 or 8 short of that found by titration and also by the nitrous-acid method of Van Slyke, which does not estimate imino-groups.

These findings raise questions of great interest, for I am convinced that our analysis for lysine is accurate to within one molecular equivalent and that no other diamino-acid is present. If one is prepared to concede certain assumptions underlying Cannan's interpretation of his data therefore it seems to me that one cannot escape the conclusion that these excess amino-groups must be those of monoamino-acids and that the molecule of β -lactoglobulin is not a simple polypeptide chain at all but an association of 8 or more probably 9 such chains held together by linkages involving carboxyl- and not amino-groups.

Before elaborating this thesis further it is necessary to consider briefly whether it is in accord with the titration data. I should expect such terminal amino-groups to have a dissociation constant of about 8, or perhaps a little higher, say 8.5, if the protons are dissociating from a molecule with net negative charge, whereas the constant for the ϵ -amino-groups of lysine would be about 10.5. Such a large

ages, therefore, especially the lateral one, might be nearly as readily hydrolysable as in diacetamide, though such a matter cannot properly be considered without reference to the specific configuration of the chains concerned. It is perhaps worth noting that an imide linkage would probably have a very weak acid dissociation constant (*c.* 10⁻⁵), but it is doubtful if this would contribute to the dissociation curves in the region that Cannan has relied on for his data.

(c) The thiol-ester linkage $R.CO.S.R'$ may be formed with the thiol-group of cysteine. This might be relatively unstable, for Neuberger (1938*b*) found that in the case of NS-diacetylcysteine the nitroprusside reaction develops very slowly and the colour remains constant for many hours, indicating a slow splitting by the ammonia employed in the reaction. If the disorientation and depolymerization of the protein molecule during denaturation should be due to, or result in, the hydrolysis of thiol-ester linkages there would be a simultaneous liberation of SH groups. It seems to me that such an explanation of what has hitherto been one of the most puzzling characteristics of denaturation is more satisfying than that based on the old supposition that the SH groups are produced by the actual cleavage of the S-S linkage, which would require a powerful reducing field.

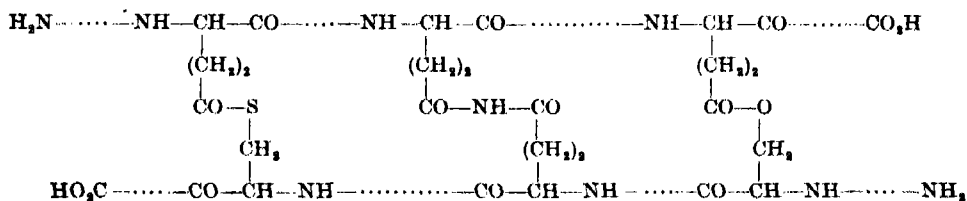


FIGURE 2. Examples of lateral thiol-ester, imide and ester cross-linkages.

Whatever types of non-peptide linkage are present in β -lactoglobulin it will be clear that if these are formed from groups belonging to terminal residues the concept of one long chain molecule is preserved (figure 1), the novel point being that the normal succession of residues is intercepted at seven or eight different positions, the intervals between which need not necessarily be the same, though random distribution is unlikely. Cross-bridges (figure 2) would give eight or nine separate chains with terminal groups free, while side-chain bridges of ester and thiol-ester linkages would give a configuration resembling a comb, the teeth of which might vary in length from one to a large number of residues (figure 3 *a, b*). In all these systems the constituent chains might not necessarily contain the same number or sequence of residues, and therein lies a possible reason why the data given in table 3 (column 5) fail to show the Bergmann-Niemann stoichiometric relationships. I shall refer to this matter again later; meanwhile I might mention here that the proposed new structure requires that on hydrolysis the protein should give seven or eight more carboxyl- than amino-groups, which might at first sight appear to contradict the observation of Hotchkiss (1939) that on both enzymic and acid hydrolysis these groups are produced in equivalent amount. It is perhaps worth mentioning therefore that the molecule of β -lactoglobulin contains about

370 peptide linkages, so that the difference to be observed would be of the order of only 2 %, probably well within the limit of error of the methods used to estimate these groups.

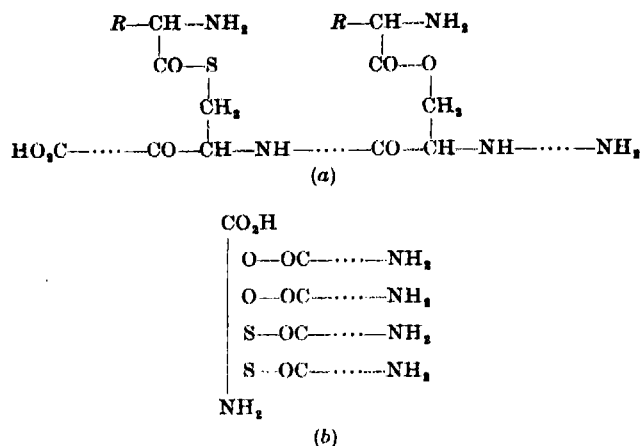


FIGURE 3. Examples of systems with branch-chain linkages.

THE STRUCTURE OF EGG ALBUMIN

Bergmann & Niemann (1937) reviewed the analytical data for this protein given in the literature, and from an inspection of the ratios exhibited by the number of g.mol. of the various residues (compare table 5, column 2) deduced (incorrectly, see Taylor 1937) that the lowest common multiple of the residue frequencies was 288 ($2^5 \times 3^2$). The number of residues in the smallest unit was therefore 288, and as the weight of the mean residue appeared to be 124 they concluded that the minimum molecular weight of the protein was 35,700. Our own analytical data, together with some recent determinations of Kassell & Brand (1938), are given in table 5. Taking 111.4 as the mean residue weight (table 1) the number of g.mol. of the mean residue will be 897, and from this the frequencies given in column 3 have been calculated. The smoothed numbers are not all factors of 2 and 3. The minimum molecular weight has accordingly been calculated from the analytical data and is close on 43,000 (column 6), in agreement with recent values obtained from sedimentation velocity, diffusion and osmotic-pressure measurements (Svedberg & Pedersen 1940; Adair 1937). The total number of residues can be taken as 384 ($2^7 \times 3$), but the residue numbers given in column 5 are not all factors of this, or factorizable by 2 and 3. Our experimental data therefore do not support the Bergmann-Niemann hypothesis.

That this should indeed be the case is to be expected from the recent work of Longworth, Cannan & MacInnes (1940), who have shown by electrophoresis that the crystalline protein contains two components. In a private communication Professor R. K. Cannan has informed me that the amount of the second component has varied widely from sample to sample (10–77 %, but generally close to the lower

limit), and that for any given sample the relative proportion of the two components is not changed after removal, by acid or heat denaturation, of one-half of the total protein present. These latter observations, taken in conjunction with the fact that previous workers have failed to effect any apparent fractionation of the protein by repeated recrystallization, would suggest that the two components are somewhat similar in composition and carry nearly the same net ionic charge. The lack of homogeneity in crystalline egg albumin therefore does not introduce a situation that differs markedly from that which I have deduced for β -lactoglobulin, and as our amino-acid analysis and the titration data of Cannan, Kibrick & Palmer (1941) were obtained from samples of material made under identical conditions a comparison of ionizable groups may still be of interest.

TABLE 5. EGG ALBUMIN

- (1) Percentage of amino-acid given on hydrolysis. The values for cystine and methionine are from Kassell & Brand (1938); the remainder are from unpublished analyses.
- (2) Number of g.mol. $\times 10^{-5}$ per g. protein, calculated from the data in (1).
- (3) Frequencies calculated from the number of g.mol. $\times 10^{-5}$ per g. protein (897) of the mean residue and the data in (2).
- (4) Minimal mol. wt. calculated from the data in (1).
- (5) Assumed number of residues.
- (6) Calculated mol. wt. of protein.

amino-acid:	(1)	(2)	(3)	(4)	(5)	(6)
arginine	5.63	32.34	27.7	3,092	14	43,280
histidine	1.45	9.35	96.0	10,070	4	42,800
lysine	5.06	34.63	25.9	2,886	15	43,310
tyrosine	4.10	22.64	39.6	4,417	10	44,170
tryptophan	1.32	6.47	139	15,460	3	46,400
cystine	1.79	7.45	120	13,420	3	40,270
methionine	5.10	34.8	26.2	2,925	15	43,900
glutamic acid	16.1	110.9	8.1	902	48	43,300
aspartic acid	8.1	61.2	14.6	1,638	27	44,100
amide ammonia	1.23	72.2	12.4	1,395	31	43,250

As in the case of β -lactoglobulin the estimate of amino-groups computed from the titration curves is in good agreement with that obtained from a determination of the free amino-N. In other respects, however, the interpretation of the acid-base binding data of egg albumin is not so straightforward. This protein contains one molecule of phosphoric acid, and such a group may be expected to contribute two equivalents to the titration curves; one of these will be submerged in the carboxyl region close to pH 2 and the other in the imidazole region at about pH 7. According to Neuberg (1938a) the constituent polysaccharide, of molecular weight about 1,250, contains 4 mols. of mannose, 2 mols. of glucosamine and an unknown basic component. We have shown that in the intact protein the amino-groups of the glucosamine are acetylated, so that these cannot contribute to the titration curves. Neuberg's own investigations suggest that the unknown basic component may contain two if not three amino-groups, but no indication was

obtained as to whether these were free, in peptide linkage or acetylated; nor was an estimate of carboxyl-groups obtained.

TABLE 6. ESTIMATES OF IONIZABLE GROUPS IN THRICE RECRYSTALLIZED EGG ALBUMIN

(Molecular weight taken as 43,000)

from data of Cannan <i>et al.</i> (1941)		from amino-acid analysis of protein		
ionizable groups	equiv. per mol.	amino-acid, etc.	ionizable groups	equiv. per mol.
guanidine	14	arginino	guanidine	14
imidazole	5	histidino	imidazole	4
amino	21	phosphoric acid		1
		lysine	ϵ -amino	15
(amino-N by Van Slyke's method)	22-23	aspartic acid	27	
		glutamic acid	48	
			75	
		ammonia	31	
			44	
		phosphoric acid	1	
free acid groups	49		free acid groups	45

I can thus do no more than surmise that the intact protein contains at least four amino-groups in excess of those attributable to lysine (hydroxylysine is absent) and not more than four carboxyl-groups in excess of those attributable to the dicarboxylic acids, leading to the conclusion that the molecule (or rather two molecules treated as one) contains at least four component polypeptide chains, all of which must have terminal free amino-groups and some must have free terminal carboxyl-groups. The data given in table 5 (column 5) show that the chains cannot all be of the same length or, alternately, that they must differ in composition. Such a deduction must be regarded as provisional only, for the titration curves of this protein do not exhibit any clear indication of amino-groups with dissociation constants of 8.5 or lower.

In spite of the rather meagre evidence that I can bring forward at the present time from purely chemical data it is, nevertheless, of great interest to note that my deduction of at least four component peptide chains is in keeping with recent X-ray evidence. Astbury, Dickinson & Bailey (1935) have shown experimentally that the individual chains in denatured egg-white must be shorter than those in denatured seed globulins (e.g. edestin—see below), and Astbury (1936, 1941) has also concluded from X-ray and Gorter's monolayer measurements that the native egg albumin molecule consists of four superposed laminae. Such a suggestion is clearly compatible with the idea of four component peptide chains cemented in some way by means of the polysaccharide and perhaps also by cross-linkages. It is clearly unwise to stress the analogy at the present time, for I am not yet in a position to bring forward conclusive chemical evidence for the presence of several

terminal amino-groups in egg albumin or in β -lactoglobulin. Fortunately, in the case of insulin I am under no such disability.

THE STRUCTURE OF INSULIN

The determinations of the basic amino-acids of this protein recorded in the literature were made by the distribution method of Van Slyke, which is unsatisfactory for several reasons. For the discussions which follow it was necessary to know fairly accurately the value for lysine, and we have accordingly redetermined the three bases by a modification of the small-scale methods of Block and Tristram which entailed a preliminary removal of the cystine as the cuprous mercaptide of cysteine. Analytical data, together with those for cystine and tyrosine, are given in table 7 (column 1): the value for histidine is probably accurate to 2 % and the

TABLE 7. INSULIN

- (1) Percentage of amino-acid given on hydrolysis. The value for cystine is from Miller & du Vigneaud (1937) and for tyrosine from Harington & Neuburger (1936).
- (2) Number of g.mol. $\times 10^{-5}$ per g. protein, calculated from the data in (1).
- (3) Frequencies calculated from the number of g.mol. $\times 10^{-5}$ per g. protein (811) of the mean residue and the data in (2).
- (4) Minimal mol. wt. calculated from the data in (1).
- (5) Assumed number of residues.
- (6) Calculated minimum mol. wt. of protein.

	(1)	(2)	(3)	(4)	(5)	(6)
arginine	3.05	17.45	46.4	5,719	6	34,320
histidine	10.7	69.0	11.8	1,449	24	34,800
lysine	1.26	8.59	94.4	11,640	3	34,930
cystine	12.5	52.0	15.6	1,922	18	34,600
tyrosine	12.5	69.0	11.7	1,449	24	34,800
amide ammonia	1.65	96.8	8.4	1,033	34*	35,120

* In agreement with Harington & Mead (1936).

other values to 3 %. The results for lysine and histidine differ significantly from those hitherto recorded, and the high value of the latter is noteworthy. I have purposely omitted the recent determinations of serine and threonine by Nicolet & Shinn (1941), as I am uncertain of the extent to which these two amino acids undergo decomposition during acid hydrolysis. The non-amino-N present in the mercaptide filtrate, determined indirectly by Van Slyke's method, was wholly accounted for by the arginine and histidine, suggesting that insulin contains very little if any proline or hydroxyproline. The calculated minimum molecular weight is 11,600, and as this is almost exactly one-third of the molecular weight found in the ultracentrifuge by Sjögren & Svedberg (1931), I shall assume that the latter is a correct value, though a higher one, 41,000, has been recorded from measurements of the diffusion constant (Svedberg & Pedersen 1940). The calculated mean residue

weight (table 1) is 123.8, and inspection of the data given in table 7 (columns 3 and 5) suggests that the protein may contain 288 ($2^5 \times 3^2$) residues, which would give a molecular weight of 35,700. Agreement is fair.

The number of free amino-groups determined under the conditions of Van Slyke's manometric method for various reaction times, together with the corresponding data of Cannan *et al.* (1942) for β -lactoglobulin, are recorded in table 8. It is to be remembered that when the protein is undergoing deamination in the reaction chamber it is not in solution, so that the number of groups that can be ascribed to lysine for short reaction periods (5–10 min.) will be definitely less than that indicated in column 3, where the values are based on the reaction rate of lysine

TABLE 8. AMINO-GROUPS OF β -LACTOGLOBULIN AND INSULIN

(Van Slyke's manometric method)

protein	(1) reaction time min.	(2) observed amino-N* equivalents	(3) contribution of ϵ -amino-N of lysine† equivalents	(4) computed free α -amino-N (2)–(3) equivalents
β -lactoglobulin	5	26.9	20.5	6.4
(per 42,000 g.)	15	35.3	27.0	8.3
$t = 22-23^\circ$	30	36.9	28.0	8.9
	60	37.5	28.0	9.5
	90	37.8	28.0	9.8
insulin	6	17.7	2.4	15.3
(per 35,000 g.)	11	20.0	2.75	17.25
$t = 20-21^\circ$	30	21.2	3.0	18.2
	60	21.7	3.0	18.7

* Computed from the data of Cannan *et al.* (1942).

† Calculated from determinations on lysine monohydrochloride, in which the α -amino-group is assumed to react quantitatively in 4 min. The ϵ -amino-group was found to react to the extent of 50, 73, 80, 93, 97 and 100 % in 2.5, 5, 6, 10, 15 and 30 min. respectively. It is to be noted that proteins are not in solution during the reaction with nitrous acid, so that the percentage reaction of the ϵ -amino-group for short time periods (3–10 min.) will be a little less than that indicated in column 3 (see text).

itself. I conclude therefore that the number of free α -amino-groups in β -lactoglobulin is eight or nine, as given by a reaction time of 15–30 min. Now I have already pointed out that in both β -lactoglobulin and egg albumin Cannan and his colleagues were able to account for all the Van Slyke non-lysine amino-N in their computations from the aqueous and formaldehyde titration curves; the aqueous titration curves moreover were reversible below pH 11, suggesting that this amino-N was not due to breakdown of the molecule. Harington & Neuburger (1936) have likewise found that no irreversible change takes place on exposure of insulin to high degrees of acidity (pH 2.5) and alkalinity (pH 11.4), the material being readily recovered later in crystalline form. I feel justified in concluding therefore that the eighteen non-lysine amino-groups indicated by the data in table 8 (column 4) represent free α -amino-groups of the native protein. The presence

of such groups has been definitely proved by Jensen & Evans' (1935) isolation of the phenylhydantoin of phenylalanine from phenylisocyanate-insulin. Insulin would thus appear to be a system of eighteen peptide chains of unlike composition, a conclusion which is in striking agreement with a property, pointed out by Bernal (1939), of the (0001) Patterson projection prepared by Crowfoot (1941) from her X-ray data. There is a distribution of eighteen peaks in the projection that suggests for the insulin molecule a structure of eighteen subunits in close-packed array. Some of these subunits must be held together by linkages, probably hydrogen bridges, that can be readily broken and remade, for Sjögren & Svedberg (1931) found that outside the stability range of pH 4.5 to about 7 reversible dissociation into products of low molecular weight occurred. It is very improbable that peptide or disulphide linkages would be involved in such changes.

ADDITIONAL EVIDENCE FOR THE STRUCTURE OF EDESTIN

The conclusion that the molecules of β -lactoglobulin, egg albumin and insulin are systems of peptide chains necessitates further consideration of my previous deduction from the purely analytical data given in table 2 that the edestin submolecule of molecular weight 50,000 is a simple peptide chain. Before discussing the evidence to be obtained from a Van Slyke determination of free amino-N there is need to recall Osborne's (1902 *a, b*) observation that when a small amount of mineral acid is added to a salt solution of this protein it is slowly and irreversibly transformed into a derivative called edestan, the average molecular weight of which has been shown by Adair & Adair (1934) to be 17,000, or about one-third that of the edestin submolecule. The formation of this interesting product of denaturation has been recently investigated in considerable detail by Bailey (1942), who has demonstrated that the fragmentation of the edestin molecule gives rise also to very small amounts of simple dialysable non-protein products, especially ammonia and tryptophan, each of which was shown to account for approximately one-third of the dialysable-N. The nitrogen content of edestan is lower than that of edestin, and the difference is inadequately accounted for by the loss of the above-mentioned simple fragment; this suggests the addition of water and by implication that edestan is formed by hydrolytic cleavage with the liberation of new end-groups. The presence of the latter was surmised in an independent manner from the alkaline shift of pH which was observed during the denaturation process.

Bailey's researches make it clear that under the conditions of the Van Slyke amino-N analysis the conversion of edestin to edestan should be extremely rapid. From what has been stated above the hydrolytic cleavage of the submolecule of molecular weight 50,000 will produce not only the three molecules of edestan, with liberation of two additional amino-groups, but also dialysable fragments. Bailey's data, allowing for the fact that they were based on an examination of what was admitted to be only a fraction of the total amount, suggest that these fragments would react to give the equivalent of about one amino-group. Allowing therefore for slow and

more deep-seated secondary reaction the data given in table 9 suggest the presence of about four free α -amino-groups, and as about three of these can be ascribed to edestan formation I infer that the submolecule of the native protein contains only one free α -amino-group and is thus one long peptide chain, in full keeping with my deduction from the analytical data given in table 2. In a private communication Professor R. K. Cannan has informed me that his preliminary estimates of the amounts of base bound by edestin at pH 9 in aqueous and in formaldehyde solutions lead to the tentative conclusion that the submolecule contains eight or nine titratable amino- and imino-groups. As eight ϵ -amino-groups of lysine are present it would appear that edestin is not a system of chains with free terminal-imino-groups.

TABLE 9. AMINO-GROUPS OF EDESTIN AND HAEMOGLOBIN

(Van Slyke's manometric method; $t = 20-21^\circ$)

protein	(1)	(2)	(3)	(4)
	reaction time min.	observed amino-N equivalents	contribution of ϵ -amino-N of lysine* equivalents	computed free α -amino-N (2) - (3) equivalents
edestin† (per 50,000 g.)	3	8	4.4	3.6
	5	9	5.6	3.4
	10	11.4	7.4	4.0
	15	12.4	7.8	4.6
	30	12.8	8.0	4.8
	60	14.4	8.0	6.4
Horse carboxy- haemoglobin‡	5	43	27	16.0
	10	49	33.5	15.5
	15	51.7	35.7	16.0
	30	53.7	37	16.7
	60	55.3	37	18.3

* See footnote †, Table 8.

† For this experiment the edestin was dissolved in aqueous sodium chloride.

‡ Lysine content 8.10 % (Vickery & Leavenworth, 1928).

Native edestin gives none of the reactions for SH groups, but Greenstein (1939) has shown that on denaturation by guanidine hydrochloride and by urea these groups are liberated in amounts corresponding to one-third and one-fourth respectively of the total cystine. One is tempted to suggest therefore that the submolecule of molecular weight 50,000 contains two residues of cystine and two of cysteine, and that the native protein is an association of six such submolecules held together by labile linkages, some of which are concerned with SH groups. In this connexion the possibility of thiol-ester linkages warrants further investigation. If the submolecule is one long peptide chain its fragmentation to give edestan must involve the hydrolysis of peptide bonds, in agreement with Bailey's experience, and as several of the residue numbers (table 2, column 5) are not divisible by three it would follow that the edestan subunits cannot all be of like composition, even

though the ultracentrifuge studies of Bailey & Philpot (vide Bailey 1942) indicate that they are of approximately equal particle weights. The small amount of ammonia produced may have originated in the cleavage, during the preliminary depolymerization stage, of labile imide bridges between submolecules; otherwise it must presumably have arisen from hydrolysis of acid-amide groups during the final fragmentation to edestan. The question must be left open for the present. Cannan *et al.* (1942) calculate that the carboxyl-groups of β -lactoglobulin and egg albumin have pK° values of 4.6 and 4.29 respectively; one might reasonably expect therefore that edestan formation within the pH range (3.9–4.9) studied by Bailey, involving as it does the cleavage of peptide, amide and possibly thiol-ester linkages, would result in an alkaline shift of pH. This is indeed what Bailey observed, but, as he himself has pointed out, the effect may well be due to the fact that the disorientation and depolymerization during the denaturation process causes a change in the magnitude of the pK values of groups such that they become titratable in the pH range concerned. It is possible that what appears to be a slightly low value recorded in table 2 (column 1) for tryptophan may be due in part to the fact that the preparation and crystallization of edestin is always accompanied by the formation of a small amount of edestan, the tryptophan content of which, according to Bailey, is only $97 \pm 2\%$ of that of edestin.

It may be noted that the submolecule of molecular weight 50,000 contains sixty-four basic groups and only fifty-four acid groups; the excess of ten basic groups is thus greater than the total for histidine, so that the isoionic point of edestin should fall at a pH above that at which all the histidine has lost its charge, i.e. at a pH about 7.5–8.5. This is fully in keeping with the early classic work of Osborne (1902 *b, c*), who recognized the basic character of this protein. He showed that edestin prepared in the usual way by saline extraction of hemp seed contained combined acid (HCl and H_2SO_4), and that in order to obtain what he called the free base it was necessary to treat the saline solution with enough dilute potassium hydroxide to give a just perceptible alkaline reaction with phenolphthalein. Bailey (1942) has found that the pH of the freshly prepared protein suspended in CO_2 -free water is usually between 5.3 and 5.5, and the sample used in our analysis contained bound HCl and H_2SO_4 equivalent to six groups per submolecule of 50,000.

GENERAL DISCUSSION

My examination of the relevant evidence adducible from amino-acid analysis, titration curves and free amino groups leads to the conclusion that the molecules of edestin, β -lactoglobulin, egg albumin and insulin are systems of peptide chains containing six, nine, about four, and eighteen components respectively. Evidence given in table 9 suggests that the haemoglobin molecule likewise contains sixteen component chains. Let me discuss the bearing that these findings have on the validity of the Bergmann-Niemann hypothesis, for we have seen that the

straight evidence for this has been somewhat conflicting. In the case of edestin the analytical data (table 2) suggest that in the smallest possible submolecule the total number of residues and also the number of residues of thirteen amino-acid and acid-amide species are severally factors of 432 ($2^4 \times 3^3$), and that the molecular weight of the submolecule is 50,000. There is also a body of independent evidence which gives support to this submolecular weight, sufficient I think to constitute a strong presumption that it is a true verdict; it might be held with some justice therefore that the molecule of this particular protein, in which all the constituent peptide chains are of like composition, is one which lends support to the Bergmann-Niemann hypothesis.

In making this assertion, however, I should be guilty of omission if I did not emphasize that evidence on two very important points is entirely lacking. In the first place no account has been taken of the 224 residues which constitute the remainder of the edestin submolecule. These represent the hydrolysis products for which methods of analysis hitherto employed do not possess the required degree of accuracy, and it is gratifying to note that several fundamentally new lines of approach to overcome these difficulties are being explored at the present time (Bergmann & Niemann 1938; Ussing 1939; Rittenberg & Foster 1940; Martin & Synge 1941 *a, b*). In the second place an evaluation of the amino-acids given on strong acid hydrolysis, however complete we may be able to make it eventually, cannot provide any direct evidence for the regular frequency of recurrence which the Bergmann-Niemann hypothesis implies; as Gordon, Martin & Synge (1941) have pointed out, such evidence can be obtained only by the isolation and identification of partial hydrolysis products of chain length sufficient to contain the

- necessary minimum number of residues, and their researches on these lines will be followed with great interest by protein chemists. Failure to obtain direct confirmatory evidence nevertheless will not necessarily imply that all the Bergmann-Niemann generalizations are false, for certain of the residues may be incorporated in the peptide chain in such a way that they exhibit more than one periodic interval of recurrence, overlapping of which would result in two or more near positions in the chain being occupied by the same residue species at more distant intervals. This may well be true with respect to the large number of arginine and dicarboxylic acid residues in edestin, whereby the close proximity of ionizable groups might create a relatively few foci of strong charge which could play some part in the hydrolysis attending edestin formation (vide Bailey 1942).

It is easy to see that under such conditions it would be difficult to prove regularities in structure by chemical analysis, and the problem becomes even more complicated if these conditions apply to proteins like β -lactoglobulin, haemoglobin and insulin in which the molecule is a system of peptide chains of varied composition, or to proteins known to be polydisperse. Clearly the analyst, whose activities at present are restricted to the simpler hydrolysis products, can contribute but little on his own, and there is need for more co-operation with the physical chemists and crystallographers, who are able to investigate the properties and structure of

larger units and of the intact protein molecule itself. The need for co-operation moreover is mutual, but progress may be slow for the art of the analyst is an exacting one. 'Meanwhile, I think that those interested in proteins would be wise to regard the Bergmann-Niemann hypothesis as still tentative and in any case as applicable only to the component peptide chains of the molecule, for much of the evidence hitherto brought forward to support it has been based on inadequate experimental data and has demonstrated nothing more than the hypnotic power of numerology.

The problems discussed in this lecture have occupied much of my spare time during the past few years, during which I have profited by many discussions with Dr W. T. Astbury, Dr Kenneth Bailey and Professor R. K. Cannan. In closing, I should like to record my deep appreciation of the wholehearted support I have received from my research assistants, Mr E. F. Williams and Mr M. W. Rees, who are responsible for most of the unpublished analytical data recorded. I have also made use of unpublished data obtained in the past by Dr G. R. Tristram, to whom my thanks are due. I have also to thank Mrs Pitt Rivers for assistance in determining the glucosamine content of deaminated egg albumin.

REFERENCES

- Adair, G. S. 1937 *Ann. Rev. Biochem.* **6**, 163.
 Adair, G. S. & Adair, M. E. 1934 *Biochem. J.* **28**, 199.
 Astbury, W. T. 1934 *Cold Spr. Harb. Symp. Quant. Biol.* **2**, 15.
 Astbury, W. T. 1936 *Nature, Lond.*, **137**, 803.
 Astbury, W. T. 1941 *Chem. and Ind.* **60**, 491.
 Astbury, W. T. 1942 *J. Chem. Soc.* p. 337.
 Astbury, W. T., Dickinson, S. & Bailey, K. 1935 *Biochem. J.* **29**, 2351.
 Ayre, C. A. 1935 Ph.D. Thesis, Univ. London.
 Bailey, K. 1937 *Biochem. J.* **31**, 1406.
 Bailey, K. 1942 *Biochem. J.* **36**, 140.
 Bergell, P. 1907 *Hoppe-Seyl. Z.* **51**, 207.
 Bergmann, M. 1938 *Chem. Rev.* **22**, 423.
 Bergmann, M. & Niemann, C. 1936 *J. Biol. Chem.* **115**, 77.
 Bergmann, M. & Niemann, C. 1937 *J. Biol. Chem.* **118**, 301.
 Bergmann, M. & Niemann, C. 1938 *J. Biol. Chem.* **122**, 577.
 Bergmann, M. & Stein, W. H. 1939 *J. Biol. Chem.* **128**, 217.
 Bernal, J. D. 1939 *Proc. Roy. Soc. A*, **170**, 75.
 Blook, R. 1939 *J. Biol. Chem.* **128**, 181.
 Brand, E. & Kassell, B. 1942 *Federation Proceedings*, Part II, **1**, no. 1.
 Burk, N. F. & Greenberg, D. M. 1930 *J. Biol. Chem.* **87**, 197.
 Cannan, R. K., Kibrick, A. & Palmer, A. H. 1941 *Ann. N.Y. Acad. Sci.* **41**, 243.
 Cannan, R. K., Palmer, A. H. & Kibrick, A. C. 1942 *J. Biol. Chem.* **142**, 803.
 Cohn, E. J. 1938 *Cold Spr. Harb. Symp. Quant. Biol.* **6**, 17.
 Crowfoot, D. 1941 *Chem. Rev.* **28**, 215.
 Dakin, H. D. 1920 *J. Biol. Chem.* **44**, 499.
 Damodaran, N. 1932 *Biochem. J.* **26**, 135.
 Gordon, A. H., Martin, A. J. P. & Synge, R. L. M. 1941 *Biochem. J.* **35**, 1369.
 Greenstein, J. P. 1939 *J. Biol. Chem.* **128**, 233.
 Harington, C. R. & Mead, T. H. 1936 *Biochem. J.* **30**, 1598.

- Harington, C. R. & Neuberger, A. 1936 *Biochem. J.* **30**, 809.
Hotchkiss, R. D. 1939 *J. Biol. Chem.* **131**, 387.
Jensen, H. & Evans, E. A. Jr. 1935 *J. Biol. Chem.* **108**, 1.
Kassell, B. & Brand, E. 1938 *J. Biol. Chem.* **125**, 145.
Longworth, L. G., Cannan, R. K. & MacInnes, D. A. 1940 *J. Amer. Chem. Soc.* **62**, 2580.
Lugg, J. W. H. 1938 *a, b Biochem. J.* **32**, 114, 775.
Macpherson, H. T. Unpublished data.
Martin, A. J. P. & Synge, K. L. M. 1941 *a, b Biochem. J.* **35**, 91, 1358.
Melville, J. 1935 *Biochem. J.* **29**, 179.
Miller, G. L. & du Vigneaud, V. 1937 *J. Biol. Chem.* **118**, 101.
Moggridge, R. C. G. & Neuberger, A. 1938 *J. Chem. Soc.* p. 745.
Neuberger, A. 1938 *a, b Biochem. J.* **32**, 1435, 1452.
Neuberger, A. 1939 *Proc. Roy. Soc. B*, **127**, 25.
Nicolet, B. H. & Shinn, L. A. 1941 *J. Amer. Chem. Soc.* **63**, 1436.
Osborne, T. B. 1902 *a, b, c J. Amer. Chem. Soc.* **24**, 28, 29, 140.
Osborne, T. B., Van Slyke, D. D., Leavenworth, C. S. & Vinograd, M. 1915 *J. Biol. Chem.* **22**, 259.
Pedersen, K. O. 1936 *Biochem. J.* **30**, 961.
Pirie, N. W. 1939 *Ann. Rep. Chem. Soc.* **36**, 352.
Rittenberg, D. & Foster, G. L. 1940 *J. Biol. Chem.* **133**, 737.
Roche, J. & Jean, G. 1934 *Bull. Soc. Chem. Biol.* **16**, 757.
Sharp, J. 1939 *Biochem. J.* **33**, 679.
Sjögron, B. & Svedberg, T. 1931 *J. Amer. Chem. Soc.* **53**, 2657.
Svedberg, T. 1938 *Industr. Engng Chem.* (Anal. ed.), **10**, 113.
Svedberg, T. & Pedersen, K. O. 1940 *The ultracentrifuge*. Oxford: Clarendon Press.
Taylor, T. W. J. 1937 *Ann. Rep. Chem. Soc.* **34**, 302.
Ussing, H. H. 1939 *Nature, Lond.*, **144**, 977.
Van Slyke, D. D. & Birchard, F. J. 1913 *J. Biol. Chem.* **16**, 539.
Van Slyke, D. D., Hiller, A. & MacFadyen, D. A. 1941 *J. Biol. Chem.* **141**, 681.
Vickery, H. B. 1940 *J. Biol. Chem.* **132**, 325.
Vickery, H. B. & Block, R. 1931 *J. Biol. Chem.* **93**, 105.
Vickery, H. B. & Leavenworth, C. S. 1928 *J. Biol. Chem.* **79**, 377.
Vickery, H. B. & White, A. 1933 *J. Biol. Chem.* **103**, 413.

Evolution in plants by kaleidoscopic mutation

By J. C. WILLIS, F.R.S.

(Received 16 April 1942)

In *The course of evolution* I have shown that there is a great probability that evolution proceeded in the direction from family to variety, i.e. in the opposite way to what is indicated by the selection theory, and that its progress was apparently by mutations that might be of any rank. The present work, begun about 1906, goes a step further, trying to show that these mutations happened in what may be called a kaleidoscopic way.

Work upon endemism for the last 40 years has shown that in general endemics are young beginners as species, first formed in the places where they now exist, and that they are neither relics nor refugees from elsewhere. It follows that their characters must have come from whatever widely distributed species of their genera are now found in the same country. Taking the numerous genera in Ceylon that show each one 'wide' and one endemic, one finds that the characters that distinguish these from one another are commonly very clearly marked and are not infrequently the same characters that mark the division of the genus into sections, or even into subgenera. But the other species of these sections are not found in Ceylon, and are usually endemic at some place a long way off. It therefore follows that the Ceylon endemics received these characters *independently, from ancestors that did not themselves show them*, by some process that reminds one of the action of a kaleidoscope. The immediate ancestor of any endemic must have been carrying *both* characters of any contrasted pair, or the factors for them, though itself only visibly showing one. Other illustrations are given, and once the conception is grasped, are evidently not far to seek.

The ideas here brought up have been gradually growing in my mind since 1907. In a paper of that date (1907*a*), in others later, and in *Age and area* (1922), many conclusions were very strongly opposed to what is usually known as Darwinism, the theory that evolution went on by the natural selection of chance variations. Further objections to it have since been published, especially in *The course of evolution* (1940), where I have endeavoured to show, with abundant evidence, that this course was in a direction opposed to that demanded by Darwinism, and was from family through genus down to species, and not in the reverse way. The evidence was mainly in the form of crucial test cases, of which 1, 2, 7, 9, 10, 12, 15, 20, 24, 26, 27 and 28 seem to the writer among the most convincing. Little or no attempt has been made, so far as he is aware, to refute their evidence. Still more were to have followed in a book upon geographical distribution, of which I have made a special study for 40 years, but in view of the present difficult situation it seems desirable to put upon record some of the new conceptions that I had hoped there to bring forward.

'It is possible, if not probable, that a group of allied species represents so many more or less stable positions of equilibrium in cell division' (Willis 1907*a*). It was already clear to me that evolution by selection could not have produced directly, by its own action, the structural changes that mark that evolution. My own view then was that some rearrangement occurred in the nucleus, at the time when a new species was about to be born, just as a rearrangement occurs when a kaleidoscope is slightly turned. The change probably occurs under some stress of conditions

acting upon the parent form (and possibly only after a certain time), as for example might be the case in the slow ascent of a mountain (Willis 1914, p. 532 and especially p. 548). If the newly rearranged nucleus were as well suited to the local conditions as its parent, it would probably succeed just as well; if not so well suited, it would probably be killed out by the action of natural selection while still only in the state of a few individuals.

This rearrangement of the nucleus would, it seemed to me, inevitably bring about at the same time, a rearrangement of the structural features of the offspring. This, apparently a necessary part of the evolution, would at the same time be the only thing to show that it had actually gone on. But though a necessary accompaniment, it was not, so far as one could see, a necessary factor in the process, which was primarily functional. At times, however, the new arrangement would doubtless be a definite improvement upon the old.

So long as the essential functions of leaves, flowers, fruit, seed, etc., are not interfered with, structural alterations probably mean very little to the life of the plant. An alternate leaf is as good as an opposite, a polypetalous flower as good as a sympetalous, a dorsal ovule-raphé as good as a ventral, and so on. None of these divergences of character offer any handle of any importance to the work of natural selection, which in any case could never produce them in the completed form in which they actually appear (cf. test case X, Willis 1940, p. 114). Evolution, in other words, involved structural changes, but was not in itself an essentially structural process. This being so, there seemed no absolute need for new structural features, and the idea that they were subsidiary, but inevitably produced by the kaleidoscopic changes that were going on, was confirmed in my mind. However different these changes may look in the kaleidoscope, the total light passing remains unaltered, as does its distribution among the colours. We shall only discuss the subject very briefly here, with a few illustrations, leaving its full treatment for a forthcoming book.

Endemism came as a perfect godsend to the supporters of natural selection, as apparently providing the necessary background of species that had been defeated in the struggle for existence, and were now dying out. But now that we know that it follows definite laws, that are concerned with increase and spread, not with dying out, this view must be abandoned, except in the cases of the few real relics that still survive, for example, within reach of the cold of the last glacial period.

In my papers from 1906 onwards, I showed that the marked endemism of the Ceylon flora was not a case of the casual and gradual extinction of the relics of an old flora, but that it obeyed certain laws, that were valid for endemics in many other places. The great majority of them proved to belong, not to small or insignificant genera that might be regarded as moribund, but to the larger genera in the country, which selectionists termed the most 'successful'. The endemics of one country belonged largely to the same genera and families as those of its neighbours (cf. Willis 1936). The endemics of a country were on the whole younger than the widely distributed species that were found there, and the smaller the area occupied

(in allied groups of species, compared only with similar groups allied to the first) the younger the species. The greater the number of widely distributed species in a country with endemics, the greater the number of the endemics, and that even genus by genus. Everything went to show that in the vast majority of cases, except in the colder northern climates, an endemic was simply a young and comparatively newly formed species that had not yet had time to spread over a wider area.

The most striking proof of all was the 'hollow curve'. The endemics of a country showed many on small areas, and there was a sharp drop to the number on the next size larger, and the drop gradually diminished less and less as one went upward (cf. Willis 1940, p. 33, for discussion and figures). Exactly the same thing occurred with the sizes of genera, the 'ones' on the average being about 38 %, the 'twos' about 13 %, the 'threes' about 7.5 %, and so on. One could not divide a curve like this into successes and failures. The only reasonable explanation, which has now been supported by abundant evidence, is that the things at the foot of the curves are the young beginners, and those at the head the oldest in the place under consideration. This explanation appears to hold for the entire flora of the world, and for many groups of animals also, in fact for all that have been tested.

From the same work I also deduced the law of size and space, that space, or area, in any circle of affinity, as before, goes with size as measured by the number of species in the genus. From these it follows that size goes with age. From these three laws—age and area, size and space, age and size—I made further deductions in 1940, that the family was as old as its first genus and species, the other genera and species being added by later mutations of specific or higher rank, without any appreciable intervention of natural selection, except as killing out anything that was not suited to the local conditions.

Putting all this together, the evidence seems almost overwhelming that the explanation of endemism just given above is the correct one, so far as we can at present see. It follows that any structural features that are exhibited in any country by its local endemics are features that must have been developed locally, and therefore, as a general rule, if there is in a genus in a given country only one species of wider distribution than that country, the rest being endemic, these characters must have come from that, *whether it shows them itself, or not*. In larger genera there may also be present some other widely distributed species from which they might have been directly inherited. We shall only give a few examples, but as will soon become evident, they could easily be multiplied many times.

There are 52 endemics in Ceylon in genera that contain only two species, one being endemic, and one widely distributed. Taking the characters in each genus that divide the two, one finds, for example:

genus	wide	endemic
<i>Ramunculus</i>	leaves much divided	leaves undivided
<i>Walsura</i>	leaves tri-foliate	leaves uni-foliate
<i>Rhamnus</i>	petals and stamens 5	petals 0, stamens 4
<i>Serpicula</i>	fr. not ribbed or tubercled	fr. strongly ribbed and tubercled

genus	wide	endemic
<i>Carallia</i>	fl. sessile; petals 7-8	fl. stalked; petals 4
<i>Alangium</i>	erect tree	subscandent shrub
<i>Mastixia</i>	petals and stamens 5	petals and stamens 4
<i>Caralluma</i>	fl. solitary, axillary	fls. umbellate, terminal
<i>Tournefortia</i>	corolla rotate, 5-lobed	corolla tubular, 4-lobed
<i>Klugia</i>	posterior angle of <i>K</i> -tube produced into a large crest	all angles of <i>K</i> -tube equally narrowly winged
<i>Cyathula</i>	fl. clusters solitary	fl. clusters in globose heads
<i>Balanophora</i>	♀ head globose	♀ head pear-shaped
<i>Putranjiva</i>	<i>K</i> of ♀ 5-6; fr. globose	<i>K</i> of ♀ 4; fr. pointed
<i>Hetaeria</i>	spike 3-5 in.	spike 4-10 in.
<i>Vanilla</i>	leaves imperfect or 0	leaves 5-7 in. long
<i>Phrynium</i>	spike lateral, high up on petiole	spike sessile on root-stock
<i>Areca</i>	stem 40-80 ft.	stem 8-12 ft.
<i>Leptaspia</i>	utricle erect, orifice terminal	utricle decurved, orifice lateral

There are also many genera with differences in size and form of leaves, or hairiness. This pattern of genus, with one species of each kind, is the most common among the genera with endemics.

These differences are just as well marked as are those given in 1940, p. 79, as first dividing the species in families with only one genus. In fact, in several cases they are the actual differences that mark the division of their genera into sections in the outline monographs in Engler (1st ed.), or even, as in *Rhamnus*, into subgenera. Other members of these divisions are found at such long distances away that it is clear that the characters that mark the sections were independently handed down to the Ceylon species, from ancestors that did not themselves show them. This table speaks for itself as showing the possibilities of direct mutation, and as explaining much apparently discontinuous distribution. Ridley's *Malayan Flora* contains at least 70 similar cases of genera with one wide and one endemic, and I have several hundreds more.

The same sort of divisions show in the Ceylon genera with one wide (*W*) and two endemics (*E*, *E* +); for example, *Alsodeia* shows distinct anthers (*W*, *E*) or connate (*E* +), *Scolopia* a green fruit (*W*, *E*) or a bright scarlet (*E* +), *Hydnocarpus* 5 stamens (*W*, *E*) or 8 (*E* +), *Sophora* 15-23 leaflets (*W*, *E*), or 7-11 (*E* +), *Chaetocarpus* a prickly capsule (*W*, *E*), or a tubercled (*E* +). *Podochilus* has linear petals (*W*), or orbicular or obovate-oblong (*E*, *E* +), *Curcuma* has the spike on a leafless peduncle (*W*, *E*) or terminating the leafy stem (*E* +). Always the behaviour is like this; as I have shown in some detail (cf. 1940, p. 74), at the first division the results are well separated, and then the later divisions fill in the gaps. The more endemics in the genus, on the whole, the more difficult does it become to separate the species.

In most monographs of genera or families, one soon finds that the characters are liable to turn up in places where one would not expect them, among groups of evidently closely related species (whether to one another or to the one showing the unexpected character). When the latter was an endemic, it was usual to regard it as a refugee from some place where the character was common, but this explanation is now no longer tenable without great and special evidence to that effect.

The Menispermaceae, for example, are divided primarily by the presence or absence of endosperm, but in both groups there are some that show the contrasted character. Both groups, also, have subdivisions with anthers opening longitudinally, and with anthers opening transversely. These characters cannot be put down to natural selection, nor can the perfection in which they are shown (1940, p. 114). Some of the subgroups have ruminant endosperm, some have not, but in each there are exceptions. But there is no need to follow this further, such experiences being part of the everyday lot of the taxonomist.

In *Symplocos* (Brand 1901), sessile flowers appear, without there being any wide to carry the character, in species endemic to such places as Colombia, Mexico, the Philippines, Formosa, Queensland, New Caledonia and the New Hebrides. A bilocular ovary appears in Cuba and Puerto Rico. A tube of united stamens is very common in America, but also appears in South India and Ceylon, in Sumatra, the Malay Peninsula, Java and Borneo, where there is no 'wide' carrying it. And so on with most of the characters, which turn up again and again in unexpected places.

Or again, take *Siparuna* (Hoffmann 1894), of which a map was given by Willis (1940, p. 34). In the aggregation of local species in Peru and Bolivia, there shown, we find characters that place their holders in the following divisions of the key, which of course uses the best and most prominent characters for the purpose:

<i>Aa</i> β I, 1	3 spp.	<i>Aa</i> β II, 2	4 spp.	<i>Aa</i> β III, 2	1 sp.	<i>Bc</i> α II, 2	1 sp.
<i>Aa</i> β I, 2	1 sp.	<i>Aa</i> β II, 3	3 spp.	<i>Ab</i> β II, 1	2 spp.	<i>Bc</i> β II, 1	1 sp.
<i>Aa</i> β II, 1	1 sp.	<i>Aa</i> β III, 1	10 spp.	<i>Bc</i> α I, 2	1 sp.	<i>Bc</i> β II, 2	1 sp.

S. guianensis, the only 'wide' occurring in this region, is placed in *Bc* β II, 2.

One might produce instances like this in very great numbers. It is the notion that endemics may be refugees that has prevented the conclusions being drawn that are drawn here.

It is a commonplace in taxonomy that single characters, such as alternate or opposite leaves, stamens in one or in two whorls, fruit loculi- or septi-cidal, ovule-raphe dorsal or ventral, and so on, cannot be relied upon to give unquestionable evidence for close relationship, though, if they were formed by gradual development from similar characters, one might expect something of the kind. One must take several characters, giving judgment upon the general evidence of them all. In a paper of 1923 and in test-case XII (1940, p. 118) I have given instances of the way in which, in Rubiaceae for example (1923, p. 624), all kinds of unexpected characters may turn up, like superior ovary, alternate leaves, imbricate calyx, anthers opening by pores, and many others that are not the usual characters of the family. But it will at once be seen, though it does not seem as yet to have been properly realized, that these characters are in no way peculiar to the Rubiaceae, but that every one of them is common somewhere else in some other family, and perhaps almost always in several other places in several different families or parts of families, or even in single genera or species. These places are usually quite

independent of one another in different parts of the scheme of classification, and it is safe to say that no system can be devised which will keep all the characters each in a line of its own. As Lindley says (Willis 1923, p. 617), 'even the sexual system of Linnaeus could not be drawn up without splitting genera, if one desired complete agreement. Smith gives 173 genera of the British flora, and no less than 43 of these, and some in every class, contain species at variance with the characters of the classes and orders.' A great deal of evidence is given in (1923) that bears upon the subject under consideration, towards which that paper formed a halfway house. If natural selection brought out characters like these, upon so many different occasions, in so many different places, and in so similar a way in each, there must have been some extraordinary urge behind it, at which we cannot even hazard a guess.

Even with the most 'important' characters, like inferior ovary, single cotyledon, parallel veining of leaves, ovary uni- or multi-locular, endosperm or none, and so on, there are always cases where they turn up unexpectedly. Consequently, each of these characters, though they are without adaptational significance, must either have been selected again and again, and always in the same manner, or must have been inherited through ancestors that were themselves visibly *carrying the contrasted character*. The same kind of thing shows all over the place, and it is very noticeable that it shows better, on the whole, the larger (more 'successful') the family. A glance at the family characters (Willis 1940, Appendix I) shows that the pairs are much fewer than the number of recognized families, whilst each family usually has one or the other of most pairs. This of course means that as we have just seen, it is almost inevitable that every character should turn up again and again. The characters of any single family are simply a hotch-potch of this list, and those that turn up least often are regarded as on the whole the most valuable as distinguishing marks. This phenomenon has always been a great stumbling block to natural selection, and again suggests the kaleidoscope. Why, if an inferior ovary is the rule for a family, should one or two want to have a superior ovary? And why should characters turn up again and again, if they have no adaptational value? And why in such a perfected form? There seems almost nothing for it but to imagine that any member of a family carries in itself all the possible characters of that family, or perhaps it is more likely that the determining factors are carried, and that, under circumstances which as yet we cannot even guess at, one or the other is brought out. Under these conditions there is hardly any limit to the possible variety. Even ten characters, which is no excessive number, allow of 3,628,800 changes, or nearly twenty times as many as all the species of flowering plants.

Let us take a few concrete cases. In his *Flora of Ceylon*, Trimen gives the condition of most leaves as regards hairiness, or none, on both sides. In *Vernonia*, for example, we find:

sp. 1 (*E*) glabrous

9 (*E*) glabrous

"

sp. 3 (*E*) glabrous or very slightly hairy

12 " glabrous or ± floccose "

"

"

sp. 2 (<i>E</i>) glabrous	sp. 11 (<i>E</i>) glabrous
slightly pubescent	woolly
13 glabrous	4 more or less hairy
finely felted	" "
5 (<i>E</i>) roughly hairy	6 (<i>E</i>) finely pubescent
" "	tomentose
7 (<i>E</i>) prickly hairy	10 (<i>E</i>) cottony or glabrous
" "	woolly felted

E=endemic. Species 8 is an introduction, and 12 only occurs elsewhere in the Nilgiris.

Selection could not produce an assortment like this, with no adaptational values, nor could it possibly make the whole of a surface alike (Willis 1940, p. 114), especially in cases like 6, 10, 11, with the two sides different. There could not have been Ceylon ancestors for them, nor, if the marks *E* were removed, could anyone say which were ancestors, or which descendants. Hundreds of similar cases could be brought up, with these or other characters.

Opposite and alternate leaves are intermingled all over the Compositae, in not very unequal numbers—a fact which goes to show that from an adaptational point of view it does not matter which is taken, so that it would be a very remarkable phenomenon indeed if natural selection, with no particular urge behind it, should choose sometimes one and sometimes the other, on hundreds of occasions in the same family. In the genera numbered 401–410 in Hoffmann 1894 we find:

401 opp. or alt.	403 alt.	405 alt. or opp.	407 upper alt.	409 opp.
402 opp.	404 alt.	406 opp.	408 radical	410 opp.

The variation in other characters in these same genera is different in its incidence. The common receptacle is long cone-shaped in 401, flat in 402 and 403, and so on. Each character differs in a different way over the ten, but each character turns up *more than once*, a fact which has been put down to the vagaries of selection, though there is nothing to show that any one character of a pair is superior to the other.

A pappus of two kinds of components (e.g. hairs and scales, an extremely difficult if not impossible thing for selection to produce) turns up again and again, usually intermingled with genera that have no pappus. It is asking too much of credulity to ask one to believe that these phenomena are due to selection.

Or let us take a small group from the key in Hoffmann (1894), e.g. the group B in Anthemideae-Chrysantheminae, nos. 557–62. Other groups show similar phenomena. Seven contrasted pairs of characters are used, which we may call *Aa*, etc. (+ means that neither shows):

557	<i>Nananthea</i>	A	B	C	+	e	F	G	1 Corsica
558	<i>Ischnea</i>	a	B	c	D	e	F	G	1 New Guinea
559	<i>Ceratogyne</i>	a	B	C	d	E	F	G	1 W. temp. Australia
560	<i>Dimorphocoma</i>	a	B	c	+	e	f	G	1 C. Australia
561	<i>Plagiocheilus</i>	a	b	C	d	e	F	G	7 S. America
562	<i>Polygyne</i>	a	b	C	d	e	F	g	1 Chile

Aa disk flowers fertile—infertile.

Ee fruit winged—not.

Bb disk flowers ligulate—two-lipped.

Ff pappus none—present.

Cc flower 4-merous—5-merous.

Gg leaves alternate—opposite.

Dd fruit angular—compressed.

In another group, nos. 479–84, where only four characters are used, these show *ABCD*, *ABcd*, *ABcD*, *AbCD*, *AbcD*, *abCD*. This same kind of kaleidoscopic exhibition of characters shows in innumerable places, and at once explains why it is practically impossible to devise any original parent for such a group, so long as one adheres to the idea that a character must be derived, even though modified, from a parental form that showed it or did not show the contrasting character. So long as this is adhered to, so long must the Darwinian direction of evolution remain. But I have shown (1940) that the evidence is almost completely against this. There seems nothing for it but to admit that any parent must be carrying in itself *both* of the contrasting characters shown in any pair, and that one or the other, or at times (as in the group of *Compositae* 401–10 above) both, in different places on the same plant, may appear, presumably according to some law which as yet we do not comprehend. No arithmetical clue has so far appeared, but it is probable that it might be found with much hard work.

The genera in these little groups, here and elsewhere, are placed there by the best taxonomists, on the ground of general close resemblance, but very commonly, as in 557–562 above, there is no proper geographical continuity among them; structural resemblance has been considered as of more importance. This is a subject for discussion in my book, and cannot be gone into in this place. Five of these six genera have *a* in common, four *B*, four *C*, three out of four *d*, five *e*, five *F*, and five *G*. There is a good majority shown for one or other of each character-pair, though not by the same genera in each case. But their geographical distribution shows that they can have no parent in their own group; if they have a common parent, it must be sought elsewhere.

The analogy—perhaps, or probably, it may be more—between this work and chemical transformation hardly needs to be pointed out. Personally, I have little doubt that in some way or other the latter is at the bottom of the changes that take place (taking chemistry in the broadest physical sense). It may be worth while to call attention to the criticism of the chemistry of evolution by Maclaren (1877), described by me in 1940, p. 8, a criticism that was met, if I remember my reading aright, by saying that the chemistry of plants was an unimportant detail so far as evolution was concerned! One of my great hopes is to see economic botany studied upon the lines indicated in 1940, p. 9, by the acceptance of the view that the parents of a genus of economic value may still be living, so that their chemistry may be studied.

Other analogies are also suggested by this work, and it looks as if, as I suggested in 1940 (foot of p. 50, pp. 175, 178, 191), evolution was based upon some mathematical conception, and that it is following out some definite plan. As Yule and Willis said (1922), 'in general the manner in which evolution has unfolded itself has been relatively little affected by the various vital and other factors, these only causing deviations this way and that from the dominant plan'.

The word 'plants' has been deliberately put into the title above to show that I have only worked out its application to them, but from the way in which so many

groups of animals give curve figures like those of plants (cf. Willis 1922, p. 200, and curve of Chrysomelidae on p. 242 and 1940, p. 37), I fancy that most of the above statements will apply to animals also, perhaps with modifications.

Incidentally, one thing is very clear, that if work is to be carried on upon the lines briefly indicated here, one thing that must be done before publication begins is to settle upon a definite nomenclature for each pair—*A*, for example, for an alternate leaf, *a* for an opposite, and so on, or there will soon be hopeless confusion.

This must suffice as an indication of what is suggested. It will be worked up, with some of its further implications, in my book upon geographical distribution, which I hope may appear in more settled times. It takes us one stage beyond that described by me in 1940, and it looks as if the next stage must fall to the biochemists or to the geneticists. Once evolution is properly comprehended, it would seem probable that it can be set to work to the immense benefit of mankind, as the results obtained with colchicine give some slight indication of possible future progress.

REFERENCES

- Brand, A. 1901 Symplocaceae, in Engler, *Das Pflanzenreich*.
 Hoffmann, O. 1894 Compositae, in Engler & Prandtl, *Die natürlichen Pflanzenfamilien*.
 MacLaren, J. J. 1877 *On some chemical difficulties of evolution*. London.
 Trimen, H. 1893–1900 *The flora of Ceylon*. London.
 Willis, J. C. 1907*a* *Ann. R. Bot. Gdns Peradeniya*, 4, 1, 17.
 Willis, J. C. 1907*b* *Ann. R. Bot. Gdns Peradeniya*, 4, 69.
 Willis, J. C. 1914 On the lack of adaptation in the Tristichaceae and Podostomaceae. *Proc. Roy. Soc. B*, 87, 532.
 Willis, J. C. 1922 *Age and area*. Cambridge.
 Willis, J. C. 1923 *Ann. Bot.* 7, 603.
 Willis, J. C. 1936 *Proc. Linn. Soc. Lond.* 148, 86.
 Willis, J. C. 1940 *The course of evolution*. Cambridge.
 Yule, G. U. 1924 *Phil. Trans. B*, 213, 21.
 Yule, G. U. & Willis, J. C. 1922 *Nature, Lond.*, 109, 177.

The production of cancer by pure hydrocarbons. VI

By G. M. BADGER, J. W. COOK, F.R.S., C. L. HEWETT,
E. L. KENNAWAY, F.R.S., (Mrs) N. M. KENNAWAY AND R. H. MARTIN

*From the Chester Beatty Research Institute, The Royal Cancer
Hospital (Free), London, S.W.3*

(Received 7 May 1942)

A series of eight 5-alkyl derivatives of 1:2-benzanthracene (methyl, ethyl, propyl, *iso*-propyl, butyl, amyl, hexyl, heptyl) was tested for cancer-producing activity. The higher members of the series showed progressive diminution of carcinogenic power upon the skin, and those having more than three carbon atoms in the side chain did not produce sarcoma. Tests upon several 10-substituted benzantracenes are recorded. The results given by several new derivatives of 3:4-benzphenanthrene confirm earlier findings in showing (1) that substitution in position 2, and to a less degree in position 1, promotes carcinogenic action, and (2) that most of these compounds are deficient in sarcoma-producing power. 1:2:3:4-Tetramethylphenanthrene, which has a weak but quite distinct property of producing tumours of the skin, is one of the simplest carcinogenic compounds known. In its molecular structure it forms a connecting link between the carcinogenic hydrocarbons of the 1:2-benzanthracene and 3:4-benzphenanthrene groups, and is also related to 1:2-dimethylchrysene which is one of the two homologues of chrysene known to be carcinogenic. 3:4:5:6-Dibenzfluorene differs from the 1:2:3:4-, 1:2:5:6- and 1:2:7:8-compounds in showing no action upon the skin.

Further tests were made of the carcinogenic power of 1:2:5:6-dibenzcarbazole in view of its considerable growth-inhibitory power. The relation between two azonaphthalenes and the diaminodinaaphthyls and dibenzcarbazoles which may be obtained from them is discussed in regard to the production of tumours of the liver. Some data are given of the occurrence of multiple and remote tumours, especially in mice receiving certain classes of compounds (2:3:4-substituted phenanthrenes, 1:2:5:6-dibenzcarbazole and 1:2:5:6-dibenzaacridine).

The usual technique employed in this institute for a number of years in testing pure compounds for cancer-producing activity consisted in repeated application of dilute solutions (usually 0.3 % in benzene) of the substance to the skin of mice. In recent years this procedure has been increasingly supplemented by subcutaneous injection of the compound, dissolved or suspended in a suitable medium. The more active carcinogenic compounds usually show similar orders of potency when assayed by both methods, but the advantage of using both techniques for the detection of activity in some of the less potent compounds was apparent from the results described in the last publication in this series (Badger, Cook, Hewett, Kennaway, Kennaway, Martin & Robinson 1940). That communication gave data for a considerable number of compounds, mostly derived from 1:2-benzanthracene, 3:4-benzphenanthrene, acridine and fluorene. In some cases addition to the food was used as a third mode of administration. Preliminary reports were given of tests still in progress, and the final results are now recorded, together with those obtained with some related compounds. For the most part the results are summarized in the tables; the discussion which follows is concerned only with some of the more noteworthy features of the results.

Tests for carcinogenic action on the skin of mice have again been carried out by repeated application of solutions in benzene (usually 0.3 %). In the tests for sarcoma-producing power sesame oil, used as a vehicle in the earlier experiments, is being replaced by glycerol, following the practice of Shear *et al.* (1940, 1941*a*, 1941*b*).

Shear, Leiter & Perrault (1941*b*) obtained negative results by the injection sub cutem of deoxycholic acid in glycerol, and in dilute NaOH, and draw attention to the contrast presented by the positive results of injections of this compound in sesame oil, reported from this Institute by Cook, Kennaway & Kennaway (1940) in a preliminary publication. These positive results, when completed, were as follows:

(1) ten stock mice, of which five lived for more than 6 months, gave spindle-celled tumours in four mice between the 351st and 421st days;

(2) spindle-celled tumours were obtained in seven out of ten C_3H mice in 208 days; one of these mice developed a spindle-celled tumour in 155 days, after receiving a total of 28 mg. deoxycholic acid in the course of six injections.*

In commenting on these experiments, Shear *et al.* (1941*b*) say, 'In view of the physiological importance of this bile-acid, it is important to ascertain to what extent the sesame oil contributed to the potency...'; and again '...although sesame oil may be of use as a co-carcinogen... it is difficult to interpret the results of tests of compounds when sesame oil is used as the vehicle'. If one is to accept this suggestion, that the induction of sarcomas in seven out of ten mice in 208 days is due to the carcinogenic or co-carcinogenic action of sesame oil, one must find certain other results also difficult to interpret, namely,

(i) Gardner (1939), whose work is quoted as the basis of objection to sesame oil, obtained only one sarcoma in seventy-seven mice receiving weekly injections of sesame oil (0.05 c.c.) 'alone or containing cholesterol or cholesterol benzoate';

(ii) in this Institute, we have tested twenty-six compounds in sesame oil upon, in all, 290 mice (280 stock, ten C_3H), without the production of a single sarcoma. In support of their suggestion that positive results with deoxycholic acid may be due to sesame oil used as a vehicle, Shear *et al.* (1941*b*) quote two statements in the literature, namely,

(a) Alessandri (1935) gave deoxycholic acid 'orally and by subcutaneous injection to mice for periods up to 100 days with negative results';

(b) Loeffler (1928) obtained negative results 'on oral administration of deoxycholic acid and cholesterol to mice and rabbits for periods up to 130 days'. Experiments of such short duration are of course quite inconclusive from the present point of view, and oral administration, whether to rabbits or to any other species, is irrelevant. Shear *et al.* (1941*b*) say that sesame oil 'is open to more serious objection than filtered lard', but they appear to overlook the fact, discovered here (Burrows, Hieger & Kennaway 1932, 1936), that lard alone may produce tumours in rats, and, much less frequently, in mice. At present we have not adequate data for a comparison of lard and sesame oil in this respect.

COMPOUNDS DERIVED FROM 1:2-BENZANTHRACENE (TABLE 1)

The carcinogenic activities of polycyclic hydrocarbons of the 1:2-benzanthracene group show a large measure of correlation with their growth-inhibitory activities. In view of the finding of Haddow & Robinson (1939) that 2':6-dimethyl-1:2-benzanthracene caused distinct inhibition of tumour growth, new tests for carcinogenic activity have been carried out with this hydrocarbon, although an

* Law (1941) injected sodium deoxycholate (? in what medium) into strain A mice and obtained (1) a fibrosarcoma in one of twenty mice in 6 months, and (2) an increase in the incidence of adenomas of the lung and in the number of adenomas per lung.

earlier experiment of 674 days' duration had not yielded skin tumours (Barry, Cook, Haslewood, Hewett, Hieger & Kennaway 1935). The new tests have likewise given negative results. The small number of mice available at the time for subcutaneous injection and the short duration of the experiment makes this negative result inconclusive.

Examination of a series of 5-alkyl derivatives of 1:2-benzanthracene, which has now been completed, has already shown how carcinogenic properties persist in this series of hydrocarbons. The production of tumours with compounds in which the substituent at position 5 was methyl, ethyl, propyl, isopropyl, butyl, amyl and hexyl has been recorded. The higher members showed progressive diminution of carcinogenic potency and also of growth inhibitory activity (Badger, Elson, Haddow, Hewett & Robinson 1942; see table 1 and figure 1). The *n*-heptyl compound, which has already been reported as giving no tumours of connective tissue, has now shown slight activity in relation to the skin (one epithelioma in twenty mice), in an experiment in which the mice lived long. It is of interest that Shear & Leiter (1940) found the 10-methyl and 10-ethyl derivatives of 1:2-benzanthracene to be potent agents in the production of sarcomas, but obtained negative results with the higher 10-alkyl derivatives (propyl, butyl and amyl compounds). These workers also found feeble activity in 10-amino-1:2-benzanthracene (two sarcomas in sixty mice). We obtained one sarcoma, and one doubtful papilloma, with 5-amino-1:2-benzanthracene.*

In regard to the other compounds mentioned in table 1 it may be noted that 10-cyano-1:2-benzanthracene has given skin tumours in two out of twenty mice, in addition to the sarcomas (four in ten mice in 333 days) previously reported; higher activity on the skin was shown by 10- β -hydroxyethyl-1:2-benzanthracene which also had given sarcomas previously; 9:10-bishydroxymethyl-1:2-benzanthracene, which had given sarcomas, was found inactive when applied to the skin, but the low solubility of this compound in benzene must be taken into account. The compound described as 10-acetyl-1:2-benzanthracene, which gave no skin tumours in an experiment of long duration (797 days), was the *meso*-acetylbenzanthracene of Cook & Hewett (1933). The position assigned to the substituent has not been rigorously established, but there is little doubt of its correctness. Shear & Leiter (1940) obtained no induced tumours by their technique with 10-cyano-1:2-benzanthracene or with 10- α -hydroxymethyl-1:2-benzanthracene. The latter differs from the compound tested by us in the position in the side chain

* A description of 5-amino-1:2-benzanthracene has not hitherto been published. The amine was prepared in this Institute by Mrs A. M. Robinson from the oxime of 5-keto-5:6:7:8-tetrahydro-1:2-benzanthracene, using the procedure whereby Schroeter (1930) transformed α -tetralone oxime into α -naphthylamine.

The oxime of 5-keto-5:6:7:8-tetrahydro-1:2-benzanthracene had m.p. 225° (decomp.). (Found: C, 82.8; H, 5.7%. $C_{18}H_{18}ON$ requires C, 82.7; H, 5.8%.) 5-Amino-1:2-benzanthracene, prepared from this oxime by the action of hydrogen chloride and acetic anhydride, formed pale greenish yellow leaflets, m.p. 173–174°. (Found: C, 88.9; H, 5.3%. $C_{18}H_{17}N$ requires C, 88.85; H, 5.4%.) Mrs Robinson also prepared the 5-hexyl and 5-heptyl derivatives of 1:2-benzanthracene (see table 1).

occupied by the hydroxyl group. The American workers found 10-formyl-1:2-benzanthracene (1:2-benzanthracene-10-aldehyde) to be active, but rather slow, in the production of sarcomas.

COMPOUNDS RELATED TO 3:4-BENZPHENANTHRENE (TABLE 2)

The results previously reported with this group of compounds revealed the pronounced influence of substituents at position 2 and (less powerfully) position 1 in promoting carcinogenic activity towards the skin of mice. The influence of groups in these positions has been further elucidated by the experiments summarized in table 2. 2-*n*-Propyl-3:4-benzphenanthrene, like the 2-methyl, 2-ethyl, and 2-*isopropyl* compounds, readily evoked skin tumours; 1-*isopropyl*-3:4-benzphenanthrene was inactive in a test of moderately long duration with 10 C₃H mice, although the 1-methyl compound had given epitheliomas in plenty. As in the 1:2-benzanthracene series, introduction of oxygen into the molecule is not necessarily attended by loss of activity. 2-Acetyl-3:4-benzphenanthrene readily gave skin tumours, and also a sarcoma; and one sarcoma, but no tumours of the skin, were obtained at the sites of application of 2-formyl-3:4-benzphenanthrene (3:4-benzphenanthraldehyde).

A notable feature of this series of compounds has been the marked superiority of the skin-painting technique over subcutaneous injection for the production of tumours. It would be of interest to determine the persistence of these compounds at the site of injection in comparison with the behaviour of the active compounds of the benzantracene series, many of which are known to persist in the subcutaneous tissues for a long time. In our experiments six compounds (3:4-benzphenanthrene and its 1-methyl, 2-*n*-propyl-, 2-*isopropyl*-, 7-methyl-, and 8-methyl-derivatives) which readily yielded skin tumours have failed to give any tumours in mice at the site of injection. Shear & Leiter (1941*a*) have recently reported failure to induce sarcomas in mice with 3:4-benzphenanthrene and five related compounds, four of which were derivatives containing substituents in position 2 (three compounds) or 1 (one compound). Sarcomas were obtained, however, by subcutaneous injection of 2-methyl and 6-methyl-3:4-benzphenanthrene (Badger *et al.* 1940); the activity of the former compound is confirmed by Andervont & Shimkin (1940).

Consideration of the molecular structures of the potent carcinogenic hydrocarbons of the 3:4-benzphenanthrene and 1:2-benzanthracene series led to the synthesis for purposes of test of 1:2-dimethylchrysene and 1:2:3:4-tetramethylphenanthrene (Hewett 1940; Hewett & Martin 1940). The production of epitheliomas (six in twenty mice) with 1:2-dimethylchrysene has been reported in our earlier communication. Like many of the 3:4-benzphenanthrene homologues named above this proved inactive in sarcoma production. It gave no tumours when administered with the food. 1:2:3:4-Tetramethylphenanthrene, which likewise produced no sarcomas, has been found to have feeble but definite carcino-

TABLE 1. DERIVATIVES OF 1:2-BENZANTHRACENE

* compound, and method of administration	initial number of mice	mice alive after		day of death of last mouse	tumours				other tumours	
		6 mths.	1 yr.		skin		epithelioma	papilloma		sarcoma
2':6-dimethyl-	A	17 C ₃ H	4	2	380	nil	nil	—	one mammary, female (124)	
	BG	4 C ₃ H	4	0	147	—	—	nil	nil	
6:9:10-trimethyl- (in addition to series in Badger <i>et al.</i> 1940)	A	10 stock	0	0	172	5 (152, 169, 170, 170, 172), two of these with paps	1 (152)	—	lung adenoma (169)	
5- <i>n</i> -hexyl-	BS	10 stock	9	3	406	—	—	nil	nil	
5- <i>n</i> -heptyl-	A	20 stock	18	9	706	1 (574) and pap.	—	—	large lung adenoma (701); two lung adenomas (574)	
									nil	
9:10-bis(hydroxymethyl)- 0.3 % in benzene (about ‡ dissolved)	A	10 C ₃ H	8	7	533	nil	nil	—	—	
5-amino-	A	10 C ₃ H	9	3	416	nil	? early pap. (416)	—	—	
	BS	10 C ₃ H	6	0	332	—	—	1 (206)	peculiar tumour, like solid mammary tumour, at s. of i. in male, after 2 x 5 mg. only (56)	
									two lung adenomas (527)	
10-β-hydroxyethyl-	A	10 stock	8	7	527	2 (356, 466) one?, partly r.u. type (416)	two paps. and two seb. adenomas (527)	—	—	
10-cyano-	A	20 stock	13	11	650	1 (397)	1 (489)	—	large lung adenoma (282), lung adenoma (506), two lung adenomas (655), stomach pap. (494)	
									lung adenoma (428, 480, 700)	
10-acetyl-	A	20 stock	15	11	797	nil	nil	—	—	

TABLE 2. COMPOUNDS RELATED TO 3:4-BENZPHENANTHRENE

compound, and method of administration	initial number of mice	mice alive after		day of death of last mouse	tumours			other tumours
		6 mths.	1 yr.		skin	epithelioma	papilloma	
2-n-propyl-	A 10 stock	6	3	549	6 including one malignant seb. adenoma (325, 549), accompanied by one or more papas. or seb. adenoma in 4; both squamous and r.a. types in one	—	—	seb. adenoma of chin in one (325); seb. adenoma of chin, and pap. on leg, in one (379); multiple lung adenomas in 3
BS 10 stock	9	6	446		—	—	—	Stomach pap., lung adenoma one large and one small, carcinoma of upper end of both ureters (375); seb. adenoma (351), stomach pap. (428), lung adenoma (422 and 447), malignant growth at pylorus (422) nil
1-iso-propyl-	A 10 C ₃ H	6	3	454	nil	—	nil	lung adenoma (263)
BG 10 C ₃ H	9	2	568	—	—	—	—	lung adenoma (228); two seb. adenomas lip (250)
2-ethyl-	A 10 stock	9	—	272	2 (214, 228)	3 (200, 250)	—	extensive spindle-celled tumour subcutem over painted area (181)
2-formyl-	A 10 C ₃ H	6	—	305	nil	—	nil	—
BG 10 C ₃ H	7	3	398	—	—	—	—	mammary carcinoma at s. of i. (337)
2-acetyl-	A 10 C ₃ H	6	—	245	2 (235, 245) in one of these, 2 epithimas. and 6 papas.	2 (208, 210) one malignant (245)	1 (366)	stomach pap. (210)
1:2:3:4-tetra-methylphenanthrene	BG 10 C ₃ H	1	—	236	—	—	1 (236)	lung adenoma (147)
A 10 stock	9	5	486	1 (399)	—	4 (392, 484) one had 6 papas. (484)	—	one pap. ventral side of neck; 4 lung adenomas (484); 3 lung adenomas (399); 2 lung adenomas (430) nil
1:2-dimethylchrysene	BG 20 C ₃ H	5	2	528	—	—	—	lung adenoma (399)
C 10 stock	8	1	399	nil	—	nil	—	nil

TABLE 3. OTHER COMPOUNDS

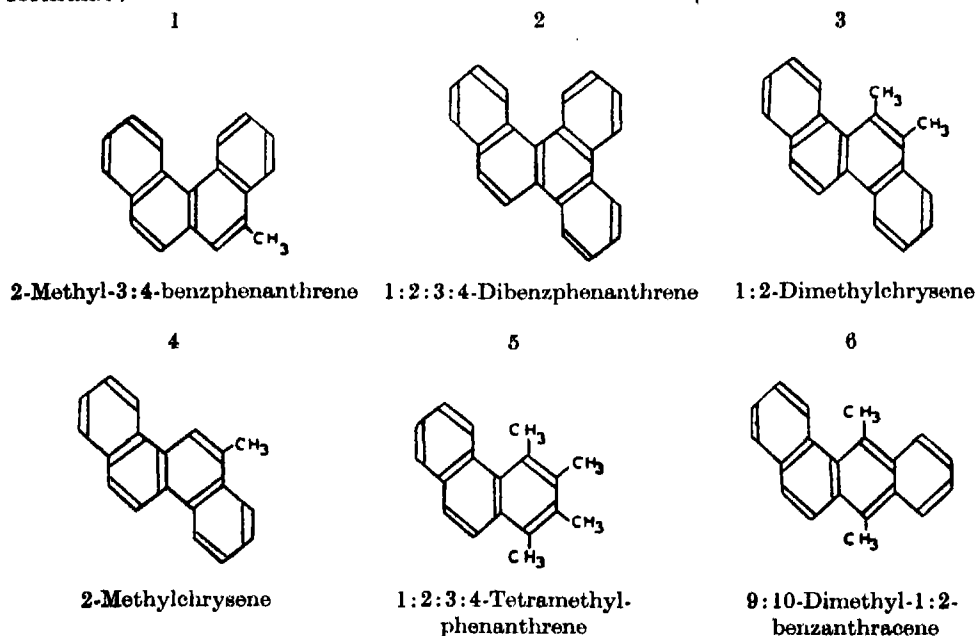
compound, and method of administration	initial number of mice	mice alive after		day of death of last	tumours				other tumours
		6 mths.	1 yr.		skin				
					epithelioma	papilloma	sarcoma		
1:2-benzfluorene	A	20 stock	10	6	600	nil	nil	—	seb. adenoma, ear and forehead (255); lung adenoma (346, 511, 595); large lung adenoma (514)
1:2:5:6-dibenzfluorenone	BS	10 stock	8	3	690	—	—	nil	lung adenoma (687)
	A	10 C ₃ H	0	0	151	nil	nil	—	nil
	BG	10 stock	1	0	200	—	—	nil	nil
3:4:5:6-dibenzfluorene	A	10 stock	6	0	354	nil	nil	—	lung adenoma (354)
	BG	10 C ₃ H	7	3	480	—	—	nil	nil
	A	10 stock	9	2	527	nil	nil	—	nil
2:3:4-trimethylfluoranthene	A	10 stock	8	6	695	nil	nil	—	lung adenoma (414)
2:2:4-trimethyl-									
1:2-dihydrofluoranthene	A	10 stock	9	2	667	nil	nil	—	nil
2:2:4-trimethyl-1:2:3:4-									
tetrahydrofluoranthene	A	10 stock	7	5	538	nil	nil	—	seb. adenoma, ear and forehead (255)
1:2:5:6-dibenzcarbazole	A	10 stock	9	—	287	5 (264, 287), one bore 3 epithemas, and one pap., one? epithema. (273) one at s. of i. (428)	1 (287)	—	had also pap. and two seb. adenomas on head (287); 3 showed lung adenomas (273, 287); 1 cholangioma (203)
	BG	10 stock	4	1	428	—	—	fibroma-like tissue at s. of i. (413) one at s. of i. (428)	in same mouse, epithema. and 2 squamous pap. of stomach, and very numerous and diffuse lung adenomas
	BG	10 C ₃ H	4	1	415	—	—	nil	in same mouse, epithema. of stomach and few small lung adenomas lung epithema. (241)

2:3:6:7-tetramethyl- anthracene; 0.3 % sus- pension in acetone; not very soluble	A	10 C ₃ H	5	nil	359	nil	nil	—	nil
	BG	10 C ₃ H	7	3	474	—	—	nil	nil
	A	10 C ₃ H	5	—	294	nil	—	—	nil
	C	11 C ₃ H	9	—	322	nil	nil	nil	nil
Di-2-naphthylethylene	A	20 stock	18	12	587	nil	nil	—	large lung adenoma (419); lung adenoma and stomach pap. (461); lung adenoma (486); 2 lung adenomas (544); large lung adenoma (617); pap. on nose and seb. adenoma on hindleg (548); ? tumour at pylorus (587)
	BG	10 stock	6	4	445	—	—	nil	lung adenoma (440)
	C	15 stock	7	5	565	nil	nil	nil	lung adenoma, and ? hepatoma (439)
	A	10 C ₃ H	9	4	464	nil			
Yellow AB	BG	10 C ₃ H	7	1	401				
	C	20 C ₃ H	16	2	403				
	A	10 C ₃ H	9	5	399				
	BG	10 C ₃ H	5	—	314				
Yellow OB	C	20 C ₃ H	13	5	491	stomach pap. (341)			

Abbreviations used in tables 1-3

Methods of administration: A = application to skin. B = injection sub cutem. BS = injection in sesame oil. BG = injection in glycerol. C = by mouth. s. of i. = site of injection. If no other statement is made, the method of administration was (1) by application to the skin twice weekly in benzene (chiefly 0.3 % solution); (2) by injection sub cutem (5 mg. in 0.2 c.c. glycerol or sesame oil) at intervals of a few weeks; and (3) by the mouth, either mixed with the food in 2 or 4 % solution in olive oil, or by stomach tube (1 mg. in 0.25 c.c. butter, or latterly margarine, to each mouse on 5 days weekly). Pap. = papilloma. Seb. adenoma = sebaceous adenoma. r.u. = rodent ulcer. Numbers in brackets show day of death of tumour-bearing mice, and pairs of such numbers within one bracket indicate the earliest and latest days of death of such mice.

genic activity towards the skin of mice. This is one of the simplest polycyclic hydrocarbons in which such activity has been found (cf. 9:10-dimethylantracene, Kennaway, Kennaway & Warren 1942). Shear & Leiter (1941a) have obtained one sarcoma in twenty mice injected with 2-methylchrysene (described as 6-methylchrysene in the American chemical literature). The structural relationships among this interesting series of carcinogenic hydrocarbons are shown by the following formulae:

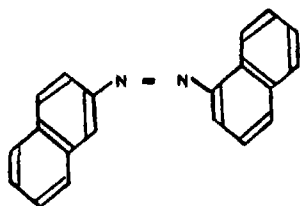


OTHER COMPOUNDS (TABLE 3)

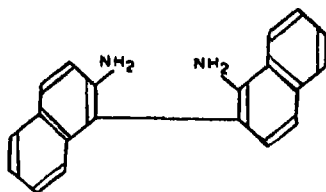
The production of skin tumours with 1:2:3:4-, 1:2:5:6- and 1:2:7:8-dibenzfluorenes has been recorded previously, but none of these gave sarcomas. 3:4:5:6-Dibenzfluorene has now been tested and gave neither skin tumours nor subcutaneous tumours. Negative results were likewise obtained with 1:2-benzfluorene and 1:2:5:6-dibenzfluorenone.

Somewhat similar in structure to the dibenzfluorenes are the dibenzcarbazoles. The 3:4:5:6-, 1:2:5:6- and 1:2:7:8-dibenzcarbazoles were shown to have carcinogenic properties by Boyland & Brues (1937). In view of the considerable growth-inhibitory power of the 1:2:5:6-compound, this substance has been further examined in respect of its carcinogenic action (table 3). This carbazole might arise *in vivo* by chemical transformation of 1:2'-azonaphthalene by reactions which have, in fact, been effected *in vitro*. Reduction of 1:2'-azonaphthalene, which is strongly inhibitory to tumour growth (Badger *et al.* 1942), gave, by rearrangement of an intermediate, a diaminodinaphthyl which is probably represented by the structure shown (Cook, Hewett, Kennaway & Kennaway 1940). Deamination of

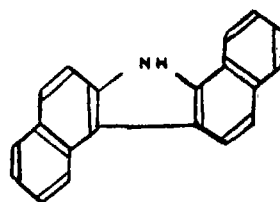
this diamine to 1:2:5:6-dibenzcarbazole is effected by prolonged boiling with hydrochloric acid (F. L. Warren, unpublished experiment):



1:2'-Azonaphthalene



1:2'-Diamino-1':2-dinaphthyl



1:2:5:6-Dibenzcarbazole

The small tendency of 1:2:5:6-dibenzcarbazole to produce tumours of the liver (some cholangiomatous tissue in one of thirty mice receiving this compound in various ways) is consistent with the view (Cook, *et al.* 1940) that the production of liver tumours by 2:2'-azonaphthalene and 2:2'-diamino-1:1'-dinaphthyl is a result of their transformation into the biologically active 3:4:5:6-dibenzcarbazole. 1:2'-Azonaphthalene and the diamine derived from it, which have little action on the liver, may well undergo similar transformation into 1:2:5:6-dibenzcarbazole, but in this case the carbazole appears to be almost without carcinogenic action on liver tissue. It will be noted that 1:2'-diamino-1':2-dinaphthyl, like the isomeric diamine, has not yielded tumours at the site of application.

The remaining compounds mentioned in table 3 have not given tumours at the site of application. The two colouring matters were tested as they are oil-soluble dyes which are included in the list of permitted colours for food in the United States. Experiments on rats with these colours are not yet complete.

MULTIPLE AND REMOTE TUMOURS

In the preceding paper of this series, attention was drawn to the need for the study of multiple and remote induced tumours in pure lines of mice in which the spontaneous incidence of tumours of the same organs is known. An attempt was made here to put this proposal into practice by substituting the C_3H strain for the stock mice used hitherto, but the C_3H mice were found to tolerate experimental treatment badly, and after the first year were much reduced in numbers. Hence we are now resuming the use of stock mice, and the results recorded here present a mixture, which is undesirable but unavoidable, of the reactions of the two kinds of mice.

Adenoma of the lung

The C_3H mice show fewer adenomas of the lung than do the stock, and this difference does not seem to be explained wholly by the longer life of the latter. Thus this form of tumour was found in thirty out of 255 stock, and in two of 191 C_3H ; large or multiple adenomas appeared in eighteen of the stock, and in none of the C_3H mice. It is perhaps significant that the two adenomas of the lung

found in C_3H mice occurred in those receiving injections of 1-isopropyl- and 2-acetyl-3:4-benzphenanthrene, i.e. compounds of a class which appears to be especially active in the induction of remote tumours (Badger *et al.* 1940, see table 7). A primary epithelioma of the lung occurred in a C_3H mouse receiving 1:2:5:6-dibenzcarbazole sub cutem. In view of the divergent behaviour of the lung in these two strains, and of the smaller numbers, it is not possible to make a detailed comparison, such as was given in the earlier paper (table 7), of the occurrence of adenomas of the lung under the influence of all the compounds tested. However, table 4 shows the comparisons which can be made if the stock mice alone are considered.

TABLE 4

class of compound	number of stock mice		percentage showing adenoma
	examined	bearing lung adenoma	
benzanthracenes	59	10	17
phenanthrenes, 1:2-dimethylchrysene, fluorenes, 1:2:5:6-dibenzcarbazole	87	24	27.6

This result agrees with those obtained previously in indicating that the benzanthracene compounds are not so effectual in producing these tumours remote from the site of application as are the phenanthrenes and some other carcinogens. One of the simplest carcinogenic compounds known, 1:2:3:4-tetramethylphenanthrene, appears to be active in this respect (see table 2). The apparent production of numerous adenomas of the lung can occur in the absence of any tumours at the site of injection (see 1:2'-diamino-1':2-dinaphthyl, table 3).

Other tumours

In the column of the accompanying tables, headed 'other tumours', are given some instances of possible remote action of carcinogenic compounds. Thus a mouse receiving 2-*n*-propyl-3:4-benzphenanthrene sub cutem showed carcinoma of the upper end of *both* ureters, a papilloma of the stomach, and a large and a small adenoma of the lung, and other mice receiving the same compound showed various sebaceous adenomas, adenomas of the lung, and innocent and malignant tumours of the stomach. These observations accord with those made upon mice receiving the closely allied compounds 2-methyl-3:4-benzphenanthrene and 1:2:3:4-dibenzphenanthrene (Badger *et al.* 1940, table 3 and p. 460). Similarly, mice receiving 1:2:5:6-dibenzcarbazole showed various tumours both of the skin, lung and stomach, and at the site of application (table 3), and in this respect resembled closely mice treated with the analogous compound 1:2:5:6-dibenzacridine (Badger *et al.* 1940, p. 459). These multiple tumours cannot be ascribed with certainty to the action of the compounds given until the spontaneous incidence of such tumours is known, but it is suggestive that all the most striking cases of multiplicity have occurred under the influence of certain classes of compounds (2:3:4-substituted phenanthrenes, 1:2:5:6-dibenzcarbazole and 1:2:5:6-dibenzacridine).

DISCUSSION

The present paper completes the reports of tests carried out during more than a decade on several hundreds of new compounds belonging to a limited number of groups of chemical compounds. These were mostly polycyclic hydrocarbons and their derivatives, the majority of them related to 1:2-benzanthracene. A recent publication of Shear & Leiter (1941*a*) is stated to 'bring to a close the investigation of the carcinogenic potency of groups of polycyclic compounds related more or less closely to 1:2-benzanthracene', carried out since 1934 under the auspices of the United States Public Health Service.

There has been an appreciable overlap in the compounds tested here and by Shear and his collaborators. Most of the compounds used by us have been synthesized here, and those used by Shear were mostly synthesized by Fieser and his school at Harvard. Thus, where the same compound was tested in both laboratories different specimens were used, often prepared by completely different synthetic methods. The mouse has been used as the test animal, but whereas we have generally used stock mice Shear's tests have been almost wholly carried out with mice of pure strains, and there have been considerable divergencies in technique in the two laboratories. In view of these circumstances the fact that there is almost complete agreement as to the activities and potencies of compounds which have been tested in both laboratories is very satisfactory.

In their final paper Shear & Leiter (1941*a*) discuss in a general way their findings, and it is unnecessary to add to their conclusions, with which we are in substantial agreement. In particular, we would emphasize the following:

(1) The compounds have, for the most part, been tested only on mice, and it may well be that a completely different order of activities would have been found in other species. Nevertheless, some of the more potent compounds have been tested in other species, with positive results.

(2) Different tissues of the same animal show a different response. This is clearly brought out by our rather extensive recent series of tests in which the same compounds have been applied to the skin and the subcutaneous tissues of mice.

(3) 'The fact that compounds of types other than the polycyclic compounds (e.g. *p*-dimethylaminoazobenzene) are also capable of inducing tumours does not detract from the specificity existing among compounds of the latter category.' The large number of compounds which have been selected for synthesis from a consideration of the structural features of known carcinogenic compounds, and found to be active, abundantly supports this view.

Shear & Leiter express disappointment that this line of inquiry has not shed light on the genesis of spontaneous tumours. In spite of this, the chemical carcinogenic compounds have provided means which are being very extensively used for the study of the production, development and behaviour of tumours. It is undoubtedly the case that the induced tumours show a very close correspondence in behaviour and character with spontaneous human tumours, and although a

study of chemical cancer-producing compounds has not disclosed substances which initiate human cancer (other than occupational) nothing has come to light to indicate that spontaneous cancer in man could not be the result of the influence of chemical compounds. It in no way follows, of course, that such chemical compounds must be of the same type as the polycyclic hydrocarbons with which we have been concerned. The carcinogenic action towards domestic rabbits of the Shope papilloma virus-protein indicates that carcinogenic properties may be found in compounds of completely different chemical types. The results obtained to date fully justify a continuation of the search in tissues for some substance which may be concerned in the initiation of malignant tumours which occur spontaneously.

This investigation has been supported by generous grants, for which we express our thanks, from the following bodies: the British Empire Cancer Campaign, the Sir Halley Stewart Trust, the International Cancer Research Foundation, the Anna Fuller Fund, and the Finney-Howell Research Foundation. We are also indebted to Dr E. de Barry Barnett for a gift of 2:3:6:7-tetramethylanthracene, to Dr S. H. Tucker and Dr J. Forrest for the fluoranthene derivatives (table 3), and to Imperial Chemical Industries, Ltd. for samples of 1:2:5:6-dibenzfluorenone and 1:2:5:6-dibenzcarbazole.

REFERENCES

- Alessandri, R. 1935 *Libro de Oro*, p. 2051. Buenos Aires.
 Andervont, H. B. & Shimkin, M. B. 1940 *J. Nat. Cancer Inst.* 1, 225.
 Badger, G. M., Cook, J. W., Hewett, C. L., Kennaway, E. L., Kennaway, N. M., Martin, R. H. & Robinson, A. M. 1940 *Proc. Roy. Soc. B*, 129, 439.
 Badger, G. M., Elson, L. A., Haddow, A., Hewett, C. L. & Robinson, A. M. 1942 *Proc. Roy. Soc. B*, 130, 255.
 Barry, G., Cook, J. W., Haslewood, G. A. D., Hewett, C. L., Hieger, I. & Kennaway, E. L. 1935 *Proc. Roy. Soc. B*, 117, 318.
 Boyland, E. & Brues, A. M. 1937 *Proc. Roy. Soc. B*, 122, 429.
 Burrows, H., Hieger, I. & Kennaway, E. L. 1932 *Amer. J. Cancer*, 16, 57.
 Burrows, H., Hieger, I. & Kennaway, E. L. 1936 *J. Path. Bact.* 43, 419.
 Cook, J. W. & Hewett, C. L. 1933 *J. Chem. Soc.* p. 1408.
 Cook, J. W., Hewett, C. L., Kennaway, E. L. & Kennaway, N. M. 1940 *Amer. J. Cancer*, 40, 82.
 Cook, J. W., Kennaway, E. L. & Kennaway, N. M. 1940 *Nature, Lond.*, 145, 627.
 Gardner, W. U. 1939 *Arch. Path.* 27, 138.
 Haddow, A. & Robinson, A. M. 1939 *Proc. Roy. Soc. B*, 127, 277.
 Hewett, C. L. 1940 *J. Chem. Soc.* p. 293.
 Hewett, C. L. & Martin, R. H. 1940 *J. Chem. Soc.* p. 1396.
 Kennaway, E. L., Kennaway, N. M. & Warren, F. L. 1942 *J. Cancer Res.* 2, 157.
 Law, L. W. 1941 *Proc. Soc. Exp. Biol., N.Y.*, 47, 37.
 Loeffler, K. 1928 *Hoppe-Seyl. Z.* 178, 186.
 Schroeter. 1930 *Ber. dtsh. chem. Ges.* 63, 1308.
 Shear, M. J. & Leiter, J. 1940 *J. Nat. Cancer Inst.* 1, 303.
 Shear, M. J. & Leiter, J. 1941a *J. Nat. Cancer Inst.* 2, 241.
 Shear, M. J., Leiter, J. & Perrault, A. 1941b *J. Nat. Cancer Inst.* 2, 99.
 Shear, M. J., Stewart, H. L. & Seligman, A. M. 1940 *J. Nat. Cancer Inst.* 1, 291.

Blood sedimentation: a study in haemo-mechanics

By R. B. WHITTINGTON, M.Sc.,
The University of Manchester

(Communicated by H. S. Raper, F.R.S.—Received 18 June 1942)

Erythrocyte sedimentation is widely used as a clinical index in certain diseases. It is a mechanical test, but its mechanics is not understood. In the present paper, sedimentation tests were carried out with constant volumes of cells at constant temperature, and the relevant physical quantities were measured at the same time. It was hoped that some insight into the operative mechanical laws might be obtained.

An equation for the maximum sedimentation velocity, in the conditions described, is proposed, from which the velocities can be forecast in terms of an independent set of measurements. The most important variable is shown to be the 'agglutination power' of the plasma, it being possible to express the maximum velocity, V_u , in the approximate formula,

$$0.759 V_u = N^{-1/2},$$

where N is the 'agglutination number' as defined.

It is shown that the effect of variations in the specific gravity of plasma and of red cells is negligible, in the experiments performed; and that while the plasma viscosity plays a relatively minor part in the sedimentation mechanics, it also acts as an index of agglutination. It follows that a clinical test, consisting of the measurement of the plasma kinematic viscosity, is possible.

INTRODUCTION

If a sample of freshly drawn blood be mixed with an anticoagulant and allowed to stand, the red cells gradually sink, leaving a clear fluid on top. *Fahraeus* (1921) found that the rate at which the cells fall varies markedly in certain pathological conditions. This sedimentation rate has since been found to be a very useful clinical index in certain diseases.

Whittington & Miller (1942) listed the physical variables to be measured during investigations of the mechanics of the test, and showed that the variable of paramount importance was the 'rouleau-aggregation' factor.

The physical mechanism of the aggregation of red blood cells appears to be imperfectly understood; but for the present purposes it will suffice to state that in the presence of an 'anticoagulant' the cells form rouleaux with contact only upon their broad surfaces, and that these rouleaux aggregate more or less loosely in networks of varying size.

When cells, washed free from their natural plasma, are suspended in saline they lose their power of aggregation, and in this state their sedimentation mechanics are considerably simplified (Whittington 1942).

In whole-blood suspensions, such as are used in the clinical test, direct measurement of the rouleau-aggregation factor presents many difficulties, but the conception of the 'effective-viscosity ratio' as a function of the agglutination pattern provides a solution to the problem.

Beginning with this conception, the present paper attempts a partial mechanical analysis of the sedimentation of pathological whole-blood suspensions, dealing with 36 % suspensions of cells at 20° C.

SYMBOLS

- F = hydrodynamic drag.
 ψ = drag coefficient.
 V = volume of cell aggregate.
 a = projected area of cell aggregate.
 d = equivalent spherical diameter of cell aggregate.
 V_t = 'terminal' sedimentation velocity.
 v = mean velocity through capillary viscometer.
 r = mean radius of capillary.
 R = mean radius of wide bore tube.
 f = viscous shearing stress on cell aggregate = $\mu_s dv/dr$.
 μ = coefficient of absolute viscosity of plasma.
 μ_s = coefficient of absolute effective viscosity of suspension.
 ν, ν_s = corresponding coefficients of kinematic viscosity ($\nu = \mu/\rho$).
 ρ = density of plasma.
 ρ_s = density of suspension.
 ρ' = density of red cells.
 ϕ = volume of cells per unit volume of suspension.
 R_e = Reynolds number ($V_t d/\nu$).
 R_N = $R_e N/d = V_t N/\nu$ ('pseudo-Reynolds number').
 $R_{N'}$ = $NR_N = V_t N/\nu_s$.
 N = the 'agglutination number' = ν/ν_s .
 ψ_N = $\psi N/d$ ('pseudo-Drag coefficient').
 G = the 'gravity factor' = $\frac{2}{3} \left(\frac{\rho' - \rho}{\rho} \right) g$.

Mechanically, the maximum sedimentation velocity is determined by the balance of the inertia and viscous forces on the cell aggregates. At this terminal velocity the acceleration is zero, so that the two forces can be equated.

The viscous drag at the terminal velocity is given by the Newton equation

$$F = \psi a v_t^2 \rho,$$

and the inertia forces are given by

$$V(\rho' - \rho)g.$$

Hence at the terminal velocity

$$\psi a v_t^2 \rho = V(\rho' - \rho)g \quad \text{or} \quad \psi = \frac{V(\rho' - \rho)g}{a \rho v_t^2}.$$

Since nothing is known of the shape of the rouleau aggregates, the ratio V/a is indeterminate; but postulating an equivalent sphere of the same drag coefficient, we may write

$$V/a = \frac{4}{3}d.$$

So the drag coefficient is expressible in terms of a single linear parameter d , since all the other terms are measurable in the equation

$$\psi = \frac{4}{3}d(\rho' - \rho)/\rho V_t^2 \quad \text{or} \quad \psi = Gd/V_t^2.$$

At the terminal velocity Reynolds's number is given by

$$Re = v_t d / \nu.$$

These are the true values of the drag coefficient and Reynolds number; but they remain indeterminate through the absence of direct measurement of d .

But assuming $N \equiv v/v_s$ (the agglutination number) to be a function of the aggregate size and shape the derived quantities (replacing d by N) may be written

$$\psi_N = GN/V_t^2; \quad R_N = V_t N / \nu,$$

and the analogous quantity $R_N = v_t N / \nu_s$.

N being measured in a suitable viscometer, ψ_N and R_N are fully determinate, and therefore they must lead to a solution for V_t in terms of the measured variables. Since N , a pure number, has replaced the linear d in the original dimensionless groups, the solution will contain a function of a linear parameter associated with N .

EXPERIMENTS

In each experiment 16 c.c. of freshly drawn blood were mixed with 4 c.c. of a 3.8 % aqueous solution of sodium citrate as anticoagulant. Great care was taken to keep the conditions of withdrawal and mixing of the blood as uniform as possible. (The samples were from patients in a Tuberculosis Sanatorium.)

The citrated blood was then centrifuged in graduated conical tubes for 45 min. at approx. 2400 r.p.m., along with a pair of haematocrit tubes to give the packed cell volume of the specimen.

The blood in one conical tube was adjusted (by the addition or removal of plasma), to give a resultant packed cell volume of 36 %. The reconstituted blood was agitated in a standardized manner and a sedimentation tube (1 ml. Westergren pipette, 2.52 mm. mean diameter) set up in a water-bath at 20° C.

Two or three viscosity measurements were carried out on the reconstituted blood during the progress of the sedimentation. All these measurements were made at 20.0° C, the viscometer being water-jacketed.

Specific gravity measurements were carried out on the reconstituted blood and on the plasma. For this purpose special specific gravity bottles of capacity approx. 1 c.c. were used. They had previously been calibrated in terms of *International critical tables* densities against a series of pure organic liquids. The bottles were always filled at 20° C.

The viscometer, essentially of the Ostwald type, though modified into a U-tube form,* and of about 1.5 c.c. capacity, was likewise calibrated at 20° in terms of *I.C.T.* kinematic viscosities against the same series of pure liquids, these being aniline, nitrobenzene, dimethylaniline and chloroform.

Finally, the kinematic viscosity of the citrated plasma was measured (at 20.0° C.).

The sedimentation curves (fall of the upper cell boundary against time) were plotted and the maximum velocities computed.

* A horizontal capillary section, with vertical limbs of wider bore.

In the tables which follow, c.g.s. units are employed throughout.

Table 1 shows the measured variables. The values of ν_s are those obtained by interpolation at 10 min. after the beginning of sedimentation, in curves of ν_s against time.

The apparent variations in red-cell density are largely due to inaccuracies in reconstituting the blood, and the mean value of 1.094 was used in calculating the derived quantities.

TABLE 1

ref. letter	V_t $\times 10^{-3}$	ν $\times 10^{-3}$	ν_s $\times 10^{-3}$	ν_s/ν	ρ	ρ_s	ρ'
A	2.00	1.872	4.5	2.4	1.031	1.056	1.101
B	1.00	1.669	4.42	2.65	1.024	1.046	1.085
C	0.833	1.716	4.46	2.60	1.026	1.048	1.087
D	3.33	1.879	4.16	2.22	1.023	1.044	1.082
E	1.50	1.867	4.31	2.31	1.027	1.047	1.083
F	0.715	1.715	4.46	2.60	1.027	1.050	1.091
H	0.354	1.617	4.7	2.9	1.026	1.051	1.095
K	3.0	1.878	4.1	2.2	1.027	1.046	1.080
M	0.333	1.640	5.4	3.3	1.021	1.056	1.118
O	1.67	1.685	3.9	2.3	1.024	1.044	1.079
P	0.306	1.663	4.91	2.95	1.027	1.054	1.102
Q	0.784	1.668	4.5	2.7	1.028	1.054	1.100
R	0.286	1.531	4.48	2.92	1.022	1.056	1.117
						mean value	1.094

TABLE 2

ref. letter	G	$\log (\psi_N \times 10^{-3})$				$\log (R_N \times 10^3)$		$\log (R_N' \times 10^3)$		$10 \times N \equiv 10\nu/\nu_s$	
		exp. 'N'		calc. 'N'		exp. 'N'	calc. 'N'	exp. 'N'	calc. 'N'	experi- mental	calc. from figure 2
		G	$\bar{G}$	G	$\bar{G}$						
A	40.0	0.620	0.658	0.634	0.672	1.649	1.663	1.269	1.297	4.17	4.31
B	44.7	1.228	1.217	1.245	1.234	1.355	1.373	0.933	0.968	3.78	3.94
C	43.4	1.382	1.384	1.381	1.383	1.272	1.271	0.857	0.856	3.85	3.85
D	45.4	0.267	0.250	0.275	0.258	1.903	1.912	1.557	1.575	4.51	4.60
E	42.7	0.915	0.924	0.896	0.905	1.542	1.523	1.178	1.141	4.33	4.15
F	42.7	1.507	1.516	1.498	1.507	1.206	1.196	0.791	0.772	3.85	3.77
H	43.4	2.077	2.079	2.076	2.078	0.878	0.876	0.416	0.412	3.45	3.44
K	42.7	0.334	0.343	0.333	0.342	1.862	1.860	1.520	1.517	4.55	4.54
M	46.7	2.106	2.076	2.158	2.128	0.789	0.841	0.270	0.374	3.03	3.41
O	44.7	0.843	0.832	0.829	0.818	1.635	1.620	1.273	1.244	4.35	4.21
P	42.7	2.189	2.198	2.187	2.196	0.795	0.793	0.325	0.321	3.39	3.37
Q	42.0	1.403	1.420	1.416	1.433	1.240	1.253	0.809	0.834	3.70	3.81
R	46.0	2.285	2.262	2.274	2.251	0.807	0.795	0.342	0.319	3.43	3.34

mean
value 43.62 = ' G '

Table 2 gives the values of the various derived quantities which are plotted in figure 1. The 'calculated N ' referred to in table 2 is derived from figure 2 which displays V_t simply as a function of N . The quantities are also worked in terms of the mean value ' G ' of the gravity factor.

Figure 1 shows that what we may term the 'pseudo-Reynolds numbers', R_N and

R_N give two straight lines when plotted logarithmically against the 'pseudo-drag coefficient'.

Clearly the lower line leads to a solution

$$V_i = \theta_1(G, N, \nu_s, L),$$

where L is a linear parameter, while the upper one gives

$$V_i = \theta_2(G, N, \nu, L).$$

Dimensional analysis shows the general form of the solutions: for if $\theta \propto G^a N^b \nu^c L^b$, dimensional homogeneity requires that

$$a = \frac{1}{3}(1+b); \quad c = \frac{1}{3}(1-2b),$$

giving solutions of the form $V_i = G^{\frac{1}{3}(1+b)} \nu^{\frac{1}{3}(1-2b)} (NL)^b$. (1)

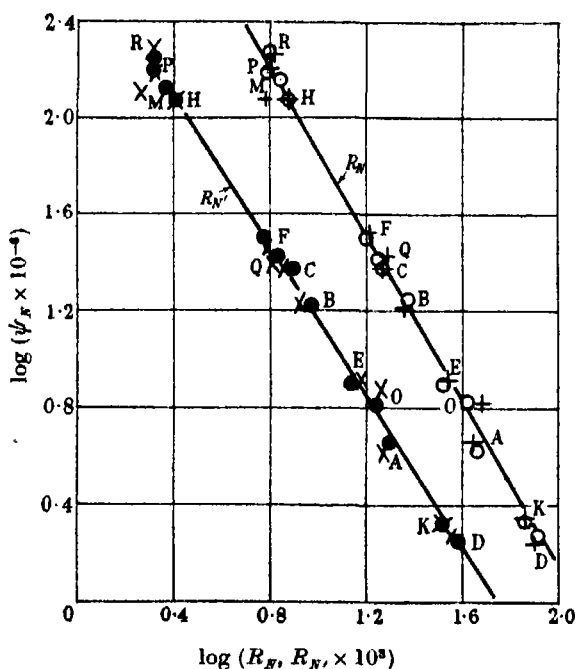


FIGURE 1

⊙ G , calc. N ; + G , exp. N ; ● $\tilde{G}$, calc. N ; × G , exp. N

The equation to the upper log line is approximately

$$\psi_N(R_N)^{7/4} = 2.3 \times 10^4,$$

leading to the solution

$$V_i = G^4 (NL)^{11/\nu^7}, \quad (2)$$

where L is about 2.6×10^{-3} cm. (b in equation (1) = 11.)

The equation to the lower log line is approximately

$$\psi_N(R_N')^{11/7} = 1.0 \times 10^4,$$

leading to the solution

$$V_i = G^{7/3} (NL)^6 / \nu_s^{11/3}, \quad (3)$$

where L is about 2.7×10^{-3} cm. (b in equation (1) = 6.)

Equations (2) and (3) are alternative representations of the same phenomena and imply that the sedimentation velocity is substantially a function of N , the agglutination number, this, as a single variable, being raised to some power between 6 and 11. The parameter L probably has a value between 2.6×10^{-2} and 2.7×10^{-2} cm., and will be discussed later.

The existence of the two equations implies the existence of a function

$$N = \theta_s(\nu),$$

and this function is tentatively displayed in figure 4.

Figure 2 shows V_t in terms of N , the most probable line having a slope of 7.7, giving

$$N^{7.7} = 0.759 V_t. \quad (4)$$

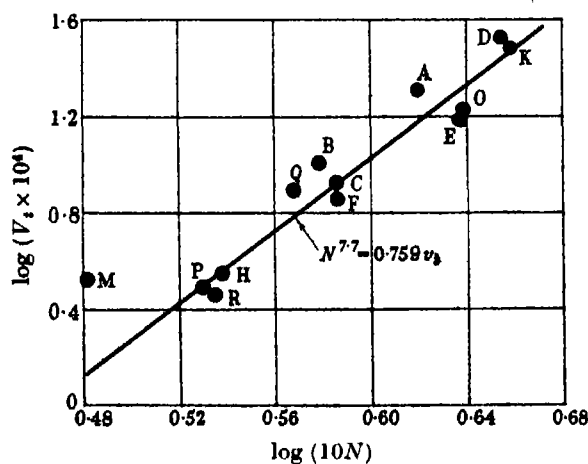
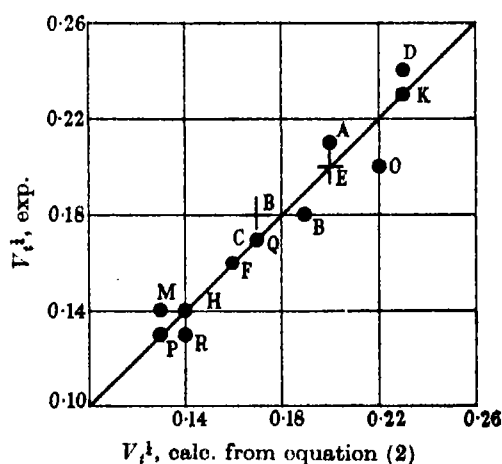


FIGURE 2

FIGURE 3. ● calc. N ; + exp. N .

Now the most difficult quantity to measure accurately is ν_s , the effective kinematic viscosity of suspension. Agglutination and packing are taking place during the measurements to varying extents. Accordingly, a calculated N was derived from figure 2, and the pseudo-Reynolds numbers and pseudo-drag coefficients were also worked in terms of this 'calculated N '.

In order to discover whether the experimental variations in G were significant or largely due to experimental errors, the derived functions were also worked in terms of $\bar{G}$, the mean value of the gravity term. The various arrangements of the derived quantities are shown in table 2 and plotted in figure 1.

Figure 3 shows values of $\sqrt{V_t}$ calculated from equation (2) plotted against the corresponding experimental values, using the value $\bar{G}$. We conclude that the experimental variations in G are of little significance.

Unfortunately, though ν is of universal application, ν_s is not, being to some extent a function of the particular viscometer employed, probably owing to the effect of the viscous shearing stresses on the elastic cell aggregates. Hence the values of N and of the parameter L in the present paper are to some extent typical of the viscometer employed.

It is desirable, therefore, to define the viscometer used, in terms of the mean viscous shearing stresses on the cell aggregates.

Let K_T be the time constant and T the time taken for the measured run, S . We have

$$v_s = T/K_T,$$

and the average velocity over the run (in the wide-bore tube of radius R)

$$= S/v_s K_T.$$

The average velocity in the capillary section (of radius ' r ')

$$= \frac{S}{v_s K_T} \left(\frac{R}{r} \right)^2.$$

The mean viscous shear, assuming $f = \mu_s dv/dr$, is thus given by

$$\bar{f} = \mu_s \left(\frac{S}{v_s K_T} \frac{R^2}{r^3} \right) = \rho_s \left(\frac{S R^2}{K_T r^3} \right). \quad (5)$$

In the viscometer used in the present experiments, K_T was 26.0×10^2 (sec./cm.)², S was 1.00 cm., R was 0.126 cm., r was 0.02 cm. Taking the mean density of a 36 % suspension as 1.050 g./c.c., we have $\bar{f} = 0.80$ dynes/sq. cm. approximately.

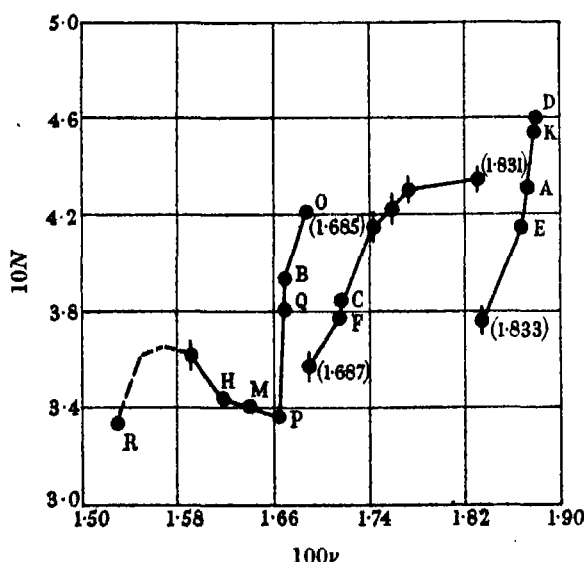


FIGURE 4

●, additional experiments

The linear parameter L evaluated from equations (2) and (3) as about 0.026–0.027 cm. is of the same order as the radius of the capillary, 0.02 cm. This value was found for the latter by measuring and weighing columns of mercury in the capillary.

We have noted the existence of a function $N = \theta_3(\nu)$, while an analogous function $V_t = \theta_4(\nu)$ should likewise exist.

The former is tentatively exhibited in figure 4, showing that plasma viscosity (ν) may be used as an agglutinative index, and that the relationship is not unique, there being a number of possible values of ν to give a single value of N ; though, on data so far available, the reverse relationship appears to be unique.

Fahraeus (1921) found that the three protein fractions, albumen, fibrinogen and globulin, made differing contributions to the agglutinating power of the plasma; and since they doubtless make differing contributions to the kinematic viscosity it is reasonable to expect that any given value of N corresponds to a number of possible combinations of the protein fractions.

CONCLUSIONS

(1) The maximum sedimentation velocity of a 36 % citrated suspension of cells at 20° C is a rough index of the agglutinating power of the plasma, and varies approximately with the 7.7th power of the agglutination number as defined above. The index of N probably varies with the proportion of cells.

(2) The maximum sedimentation velocity of any given suspension of cells at 20° C is given by the general formula

$$V_t = G^{b(1+b)} \nu^{4(1-2b)} (NL)^b,$$

where b probably varies with the proportion of cells; and the maximum sedimentation velocity of a 36 % suspension at 20° C may be approximately computed from the formula

$$V_t = G^4 (NL)^{11/\nu^7},$$

where L is a constant for the viscometer employed (0.026–0.027 cm. in the present case), and N also probably depends to some extent upon the dimensions of the viscometer. The viscometer used has been defined as exerting a mean intensity of shearing stress upon the cell aggregates of 0.80 dynes/sq. cm. Experimental variations of G appear to be of no significance.

(3) The agglutinating power of the plasma can be demonstrated as a function of the kinematic viscosity of the plasma alone. The units of kinematic viscosity employed are cm.²/sec., derived from *International Critical Tables* of absolute viscosity and density, and are therefore of general application. Ultimately, therefore, the presence of the red blood cells can be dispensed with.

The author is most grateful to the Crossley Sanatorium for laboratory facilities.

REFERENCES

- Fahraeus, R. 1921 *Acta Med. Scand.* 55, 3.
 Whittington, R. B. 1942 *Phil. Mag.* 33, 68.
 Whittington, R. B. & Miller, A. K. 1942 *Brit. J. Exp. Path.* 23, 56.

Anniversary Meeting of the Royal Society and Newton Tercentenary Celebrations, 30 November 1942

Before proceeding to the presentation of the Medals awarded for this year, it is fitting that, in accordance with custom, we should briefly recall the lives and the achievements of those whom death has removed from our Fellowship and our Foreign Membership since the last Anniversary Meeting of the Society.

His Royal Highness the DUKE OF CONNAUGHT, who died on 16 January at the advanced age of 91 years, was elected to our Fellowship under Statute 11 as long ago as 1906, so that he was by many years the senior of the members of the Royal Family who have accepted election to our Fellowship. His Royal Highness had other contacts with the advancement and the applications of scientific knowledge as President of the Royal Society of Arts, and of the Royal Colonial Institute. The latter appointment had a particular fitness in recognition of his many and great services to the British Commonwealth of Nations, as Governor-General of the Dominion of Canada and on other special missions.

EMILE PICARD (1856–1941), perpetual secretary of the Paris Academy of Sciences since 1917, had been a Foreign Member since 1909. He was one of the most famous of modern French mathematicians. He did work of fundamental importance in the theory of differential equations, the theory of algebraic functions of several variables, and the theory of surfaces and integrals associated with them; and 'Picard's theorem' is one of the classical theorems of the theory of analytic functions.

TULLIO LEVI-CIVITA (1873–1941) was elected a Foreign Member of the Society in 1930, and represented the brilliant school of mathematics in Rome. Gifted with a remarkable command of the resources of mathematical analysis and with a keen geometrical intuition, he applied his great technical skill to nearly all branches of mathematics, but it was to problems arising in applied mathematics that he devoted most of his efforts. Electrostatics, analytical dynamics, celestial mechanics, hydrodynamics, relativity, and quantum mechanics, as well as problems in engineering, all claimed his attention in turn. Mathematicians, however, associate Levi-Civita first of all with the Absolute Differential Calculus, which he helped to found with his teacher Ricci. It was in 1917 that Levi-Civita advanced the subject greatly by introducing the idea of parallel displacement. This discovery, following on the announcement of Einstein's general theory of relativity, in which the absolute differential calculus had proved a most powerful tool, provided a means of making important advances in relativity theory, culminating in the unified

theory of Weyl. Levi-Civita's numerous text-books are models of lucidity, and two of his treatises, *The Absolute Differential Calculus* and *Rational Mechanics*, are among the leading works on the subjects with which they deal.

In the EARL OF BERKELEY (1865-1942) the Society has lost a Fellow of versatility and distinction, who used his private fortune to build and equip a laboratory in which he made an outstanding contribution to the measurement of osmotic pressure. He had shown promise of a fine career in the Navy, but he left the Service when he was 22 to devote himself to research. He soon became a brilliant experimenter, and the absence of any formal training was in some ways an advantage to a man with his instinctive grasp of scientific method, as his approach to every problem was along original lines. In the great days of the Foxcombe laboratory he had a well-balanced team of workers, including Ernald Hartley, Charles Burton and M. P. Appleby, and he proved himself a fine leader with initiative and imagination. His work on osmotic pressures gave ample scope to his engineering and mathematical ability, and the agreement of the directly observed values with those calculated from vapour-pressure measurements established once for all the validity of the indirect method.

Sir WILLIAM HENRY BRAGG (1862-1942), who died on 12 March in his eightieth year, had been our President for a period so recent as 1935-40. Even more recently, during my own absence in America, he had come, from retirement, to occupy this chair as Vice-President for the two months from 20 December to 20 February last, and thus to give again to the Society's service the full measure of his wisdom and experience, up to a few weeks before a brief illness brought the end. He occupied a very special place in the scientific life, not only of the Royal Society, but of the whole nation, and, indeed, of the world at large. Without any compromise of the highest standards in his experimental and theoretical researches, he had made himself supreme as an interpreter of the results of science to those least equipped with special knowledge; so that he seemed equally happy and assured when discussing the complexities of crystal structure with fellow experts, and when presenting and enlivening the elements of science before a fascinated audience of children.

From Trinity College, Cambridge, and a high place among the Wranglers, he went in 1887 as a young man to Adelaide, to occupy a new chair of mathematics and physics. There followed a period of many years devoted to teaching and organizing, and his career as an experimental investigator, in which he attained the highest rank, was remarkable, if not unique, in the fact that it did not begin till he was 42 years of age. It was not, indeed, until he had returned to England and to the chair of physics at Leeds that, at the age of 46 years, he began, in collaboration with his son William Lawrence Bragg, the researches on crystal architecture as revealed by the reflexion of X-rays from crystals, which established his fame. They also provided the opening to the long series of masterly investiga-

tions, in which a succession of brilliant collaborators worked under his guidance and inspiration for some years at University College, London, and then, for the rest of his life, at the Royal Institution and its Davy-Faraday Research Laboratories. Here Bragg found the ideal setting and opportunity for the exercise of his distinguished gifts for research and for popular exposition, and of the hospitality and kindness which endeared him to so wide a circle.

JEAN PERRIN (1870-1942), who died in New York on 17 April, had been Professor of Physical Chemistry at the University of Paris for about forty years. He was one of the best-known physicists of his day and made valuable contributions to the knowledge of atomic physics, both by his experimental work and by his books on modern physics. In 1926 he was awarded the Nobel prize for the remarkable work he had done on the Brownian movement of small particles suspended in a liquid. By a series of newly devised methods of measurement he succeeded in making a definite estimate of the mean energy of agitation of the particles, and thence deduced the number of molecules per cubic centimetre of a gas at normal temperature and pressure. He had many interesting views on physical phenomena which are described in an attractive manner in his books, *Les atomes* and *Les éléments de la physique*. In a later publication, *Grains de matière et de lumière*, he discusses the modern theory of radiation and ionization.

Perrin also took a great interest in scientific and industrial research, and in his capacity of scientific adviser to the French government he was largely responsible for the organization of research in applied physics and the establishment of the Palais de la Découverte.

GEORGE GERALD STONEY (1863-1942), a life-long friend and colleague of the late Sir Charles A. Parsons, was world famous as one of the best-known pioneers of the steam turbine and high-speed dynamo electric machine. Educated privately and later at Trinity College, Dublin, he was one of the original staff of the steam-turbine and electrical works founded by Parsons at Newcastle-on-Tyne, and was the sole survivor of the original crew of the experimental steam yacht *Turbinia*.

From the foundation of Heaton Works in 1880, Stoney ably assisted Sir Charles in the long struggle that lay ahead to achieve recognition of the merit of the compound steam turbine, till it became the essential instrument for providing power on the largest scale. Stoney had the privilege of being an active witness of this epic of engineering history, and although he sought no publicity or public recognition of his work he became almost as well known in the world of power generation as Sir Charles himself. Content simply to support his chief to the best of his ability, Stoney contributed many papers to engineering societies and journals. He became a Fellow of the Society in 1911.

Sir JOSEPH LARMOR (1857-1942), Copley Medallist and for eleven years Secretary of the Society, was Lucasian Professor of Mathematics at Cambridge from 1903

to 1932. His distinguished contributions cover many branches of mathematical physics and geophysics. He will be especially remembered for his fundamental researches on the relations of matter, electricity and ether. These, together with the contemporary work of Lorentz, are the bridge which unites the revolutionary theories of the present century to the older physics. To Larmor modern atomic theory owes the 'Larmor precession' and the formula for the radiation of an accelerated charge, which are the basis of its most fertile developments. As an Irish mathematician he had strong attachment to the school of Hamilton, MacCullagh and FitzGerald, and he worthily carried on its great tradition.

The Society lost its senior Fellow by the death on 2 June of ANDREW RUSSELL FORSYTH (1858-1942), who had been elected at the age of twenty-seven in 1886. After graduating as Senior Wrangler in 1881, he produced in rapid succession a series of important memoirs on Theta-functions, Abel's Theorem, and Invariant-theory. In 1893 he was elected to the Sadleirian chair at Cambridge in succession to Cayley; in this position he rendered services of the greatest value to British mathematics by making known, in lectures and advanced treatises, the fruits of continental research. He was an exceptionally able administrator, and took a prominent part as a reformer in many academic movements and controversies.

Forsyth left Cambridge in 1910, and from 1913 to 1923 was Chief Professor of Mathematics in the Imperial College of Science and Technology. In 1897 he was awarded the Royal Medal: and in the course of his long life he received very many honours from Universities and foreign Academies. His devotion to the Royal Society, and to the promotion of science in fields far distant from those in which he was active himself, is shown in the bequest of his estate to the Society for the support of medical researches.

ALFRED DANIEL HALL (1864-1942) probably did more for the introduction of science into modern agriculture than any other man in our time. He possessed a many-sided ability and a remarkably wide range of interests, including Japanese prints, tulips, music, poetry and French literature; and he was a gardener of distinction. After leaving Oxford he spent a few years as schoolmaster and University Extension Lecturer. He then started the Agricultural College at Wye, Kent, where he worked out courses of instruction far ahead of any then existing. Later he took charge of Rothamsted and made important investigations on soils and crop production. Facilities, however, were very inadequate, and, realizing that more must be provided, Hall transferred his activity and his centre of interest to the Development Commission, where he played an important part in inaugurating the system of agricultural education, advice and research that has contributed so greatly to the advancement of the science and the industry.

WILLIAM HENRY YOUNG (1863-1942), Sylvester Medallist of the Society in 1928, was one of the most profound and original among British mathematicians of the

last fifty years. He was one of the leaders in the development of the modern Cambridge school of analysis, and there are few branches of the theory of functions on which he has not left his mark. The theory of sets of points, the foundations of the differential and integral calculus, the theory of Fourier series and other orthogonal developments, are full of striking theorems discovered by Young.

WILLIAM MATTHEW FLINDERS PETRIE (1853–1942) was one of the greatest figures the world has seen in the comparatively young science of Egyptology. In his own branch of that science he was pre-eminent. By his excavations in Egypt over a period of forty years he contributed more than any other single scholar in the last half-century to our knowledge of the history of the land in ancient times, and to the use of that history as the chronological yard-stick for the ancient world in general, prior to the age of classical Greece. Above all he was the founder of the scientific method of modern archaeology.

Owing to delicate health in childhood Petrie was educated privately, and largely by finding in his own way the line of his own interests—geology, surveying, chemistry and history. He first went to Egypt to survey the Great Pyramid in 1880. He returned in 1883 to start excavating for the Egypt Exploration Fund, but by 1886 he was working there on his own account, and remained for the rest of his life virtually his own master. In 1892 he was appointed to the Edwards Professorship of Egyptology at University College, London, the first chair in the subject to be founded in this country, from which he retired in 1933. He was elected F.R.S. in 1902, and knighted in 1923.

RICHARD WILLSTÄTTER (1872–1942) was an illustrious organic chemist, distinguished alike for the success that attended his attack on the most difficult problems and for the improvements that he made incidentally in the technique of investigation. His analytic and synthetic skill were shown in the determination of the constitutions of atropine and cocaine and the disclosure of their relation to *cycloheptane*. Work on aniline black and the orthoquinones opened up new vistas. Having isolated the chlorophylls, he made the surprising discovery of their magnesium content and degraded them, step by step, to porphyrins analogous to those obtainable from the blood pigments. He made important contributions to our knowledge of photosynthesis. Typical of his genius was the isolation of the anthocyanin pigments of fruits and blossoms, their recognition as glycosides and the elucidation of the structure of the latter. In later years he devoted the whole of his energies to biochemical work, for example, to the concentration and separation of the enzymes. Willstätter's premature retirement was the result of a refusal to compromise with forces and tendencies alien to his whole mentality.

EDWARD FAWCETT (1867–1942), who died suddenly at Bristol on 22 September, had a distinguished career as an anatomist. He received his medical education at the University of Edinburgh, graduating M.B., C.M. in 1889 and M.D. in 1906. After a period as Demonstrator of Anatomy at the Yorkshire College, Leeds, he

was appointed in 1893 to the Professorship of Anatomy in University College, Bristol. In spite of a heavy burden of administrative work, Fawcett found time to carry out much original research and came to be recognized as a foremost authority on the morphology and development of the mammalian skeleton, including that of man. In particular, he published in the *Journal of Anatomy* a series of papers, based on accurate wax reconstructions, in which he greatly extended our knowledge of the development of the chondro- and osteo-cranium in various species of mammals. A man of very varied interests, Fawcett in his later years became keenly interested in church architecture. He was elected a Fellow of the Society in 1923.

The death of GEORGE GERALD HENDERSON (1862-1942) has taken from us the doyen of Scottish chemists and closed a career of remarkable usefulness and influence. His thirty-five years of professorial life were spent first in the chair of chemistry in the Royal Technical College, Glasgow, and afterwards as Regius Professor in the University of that city. When he retired he left behind him two vigorous schools of chemistry, which he had consolidated through his zeal for research and his capacity to quicken enthusiasm in young men. Henderson was a brilliant lecturer, an able administrator, and a versatile scientific inquirer. Between his first paper on Dolomite and his last on the Carophyllene Series are found more than sixty valuable publications, most of which deal with the chemistry of the terpenes; but almost any type of problem—organic or inorganic, academic or technical—appealed to his seeking mind and yielded to his skill. He became a Fellow of the Society in 1916 and occupied with dignity and distinction the Presidency of the Chemical Society and of the Society of Chemical Industry.

By the death of JOHN NORMAN COLLIE (1859-1942) on 1 November chemistry has lost one of its most outstanding personalities. To his numerous friends and old students the memory of him will be one of a great organic chemist endowed with a singular vision and understanding.

Collie was characterized by a remarkable versatility, for he excelled as a mountaineer and an art critic, as well as an authority on early printing and incunabula, Chinese porcelain and lacquers, and Japanese netsukes and sword-guards. He was a true aesthete, for beauty made a strong appeal to him, whether in natural form or in the handiwork of man.

To those who knew him, Collie was gifted with an understanding of human nature which made him an ideal teacher. Although he was somewhat hesitating in his style, his lectures were singularly attractive, because there always appeared to be something of absorbing interest beyond the horizon of his description.

Award of Medals, 1942

THE COPLEY MEDAL is awarded to Sir ROBERT ROBINSON. He is recognized in all countries as one of the world's leaders in organic chemistry, and is one of the greatest and most versatile of investigators in that department of science. His researches, with a long and notable succession of pupils and collaborators, have covered a remarkably wide range of problems in this field, and his approach to these has been distinguished by brilliance in conception and a genius for the selection of methods leading to the desired solutions.

Robinson's investigations have been particularly concerned with the chemistry of natural substances, products of the life processes of plants and animals. His work has thus been a potent factor in the tendency of organic chemistry to return, in recent years, to an objective nearer to that of its origin, and to make contacts of growing intimacy and value with biochemistry, a more recent development in response to the stimulus of functional biology.

This occasion does not permit any attempt at a complete or detailed survey of all the different fields which Robinson's work has illuminated and opened to further exploration. Special mention must be made, however, of his long series of fundamental investigations on the constitution and relationships of the plant alkaloids. His theory of the biogenesis of plant products seems rather to have inspired than to have resulted from his own early and elegant synthesis of tropinone; and it has revealed an unforeseen and coherent relationship between the constitutions of different groups of alkaloids, and given a great stimulus to work on their synthesis. The work published from Robinson's laboratory has been fundamental to understanding of the isoquinoline and the indole series of alkaloids, of morphine and its allies, and of the structural formulae of strychnine and brucine, which formed the subject of his Bakerian Lecture.

Of at least equal scientific rank is the work which Robinson carried out and inspired over many years on the anthocyanin and, more recently, on the anthoxanthin pigments of plants, culminating in the synthesis of the actual colouring matters of flowers, and forming as a whole one of the most brilliant achievements in the whole range of modern organic chemistry.

Robinson's mastery of synthetic resources, and his penetrating instinct for clues to organic constitution, have been further demonstrated in a more recent approach to the synthesis of the steroids, in the production of series of compounds of interest for chemotherapy, and in notable studies of individual natural substances of a range of other types. He is, moreover, a philosopher as well as a master of experimental possibilities; and his theory of organic reactions, in the modern, electronic terms of valency bonds, has had a great influence on the development of fundamental conceptions in organic chemistry.

The Copley Medal is the highest recognition of scientific achievement in the Royal Society's gift, with no limitations of subject or nation; and the Society may well find cause for satisfaction in the knowledge that the award of its premier

Medal for achievement in organic chemistry, after an interval of many years, finds among its own Fellows a recipient of such unquestioned pre-eminence as Sir Robert Robinson.

The RUMFORD MEDAL is awarded to Dr G. M. B. DOBSON. The Rumford Medal awarded by the Royal Society was established for the recognition of important discoveries made in Europe, especially on heat or on light. These conditions appear to be met with a special fitness in the award of the medal this year to Dr G. M. B. Dobson, for his meteorological researches and discoveries. For Dobson's work has, in recent years, greatly extended knowledge of the linkage between the behaviour of ozone and cyclonic disturbances, in that complicated heat engine which is the earth's atmosphere; while light may be said to have provided the basis of measurement for Dobson's spectrographic studies of the distribution of ozone, in time and in height above the earth's surface. Light had also a major concern in earlier researches on meteors in which Dobson collaborated with Professor Lindemann (now Lord Cherwell); in these, the study of the heights between which meteors become luminescent enabled them to draw conclusions as to the density, and from these as to the temperature of the atmosphere at great heights from the earth's surface. But it is especially on Dr Dobson's own studies of the ozone of the atmosphere, continued over many years, and producing results of outstanding importance for meteorology, that the award is based.

A ROYAL MEDAL is awarded to Professor W. N. HAWORTH, a brilliant leader in organic chemistry. Haworth's great claim to distinction arises from the revolutionary change which has been produced by his own investigations, and by those of his immediate pupils, in the whole aspect of an important section of organic chemistry, dealing with the structure and the relationships of the carbohydrates. The ring structure of the simple sugars, first proposed by Tollens and supported by the work and the authority of Emil Fischer, had been generally regarded by chemists as firmly established. Detecting insecurities in the arguments which led to this formulation, Haworth developed the methylation technique, first used by Purdie and Irvine, and applied it systematically to the monosaccharides. He was thus able to show, by unequivocal methods of organic chemistry, that the accepted ring structure of these sugars was incorrect, and that, in their normal and reactive forms, they were derivatives of pyran and of furan respectively.

Later Haworth further developed his methods in application to sugars and carbohydrates of increasing complexity. By his work, and that of others who have followed his lead, the detailed structures of many disaccharides and of some trisaccharides have been established. Progress has further been made in his attack on the structural complexities of the polysaccharides, and a simple chemical method has been evolved for determining their minimal molecular weights. Professor Haworth's work, in the field which he has thus made his own, has received the high international recognition of a Nobel Prize, and will assuredly take rank as a major achievement of permanent significance in chemical history.

A ROYAL MEDAL is awarded to Dr W. W. C. TOPLEY, who is one of the most distinguished of the British bacteriologists of recent years. The most important of Topley's contributions to bacteriology and experimental medicine has been the experimental study of epidemics, which he initiated and of which the methods have been largely his own creation. Much had been learned by the statistical analysis of observational data, dealing with the origin, spread and development of natural epidemics, under conditions largely out of control. Topley conceived the idea of applying such methods to the investigation of epidemics started artificially in populations of healthy mice, kept in the laboratory under conditions which could be exactly controlled and deliberately varied. Thus the factors conducive to the rise, culmination and decline of an epidemic, to its revival or its final subsidence, could be experimentally determined. In a long series of such studies, the important results of which were reviewed in his Croonian Lecture for 1941 on *The Biology of Epidemics*, Topley had the statistical co-operation of Professor Major Greenwood, in both planning and interpretation.

On this firm basis of knowledge concerning the incidence and mortality of a naturally transmitted infection in untreated stock, the efficiency of prophylactic measures could be put to a controlled test. Under Topley's guidance and inspiration, accordingly, substantial progress had been made by his chemical colleagues towards the isolation from various species of pathogenic bacteria of highly purified and stable antigens, and the practical trial of some of these was interrupted by the outbreak of war.

Dr Topley's researches have had a great and lasting influence on the study of bacteriology, immunology and epidemiology in relation to human medicine. His recent change of the focus of his interests may be expected to give an important stimulus to advance in many cognate fields of agricultural research.

The DAVY MEDAL is awarded to Professor C. N. HINSHELWOOD for his work on the kinetics of chemical change, characterized by its pioneering quality and by the varied new lines of research which it has opened up. An experimental investigator of great skill and achievement, Hinshelwood has also enlarged the theory of the subject by able mathematical analyses and descriptions based on the concepts of collisions and of activation energy.

Hinshelwood took a leading part in the early study of homogeneous gaseous reactions. As the result of the examination of a number of bimolecular examples he was able to show, with reason, that these are confined to molecules containing few atoms, and that the actual rate is given by the product of the total collision rate and the probability of a molecule possessing the experimental energy of activation. Unimolecular reactions were found to occur with polyatomic molecules and to show more complex features. In association with Professor Lindemann (now Lord Cherwell), Hinshelwood put forward the mechanism, now accepted, whereby a reaction fundamentally dependent on collisions may nevertheless have unimolecular

kinetics. This theory he was able to verify by showing that the rates of such reactions diminish at low pressures and that the kinetics then become bimolecular.

In the field of chain reactions Hinshelwood opened up new lines of advance by studying the thermal reaction between hydrogen and oxygen. Thus he discovered, and offered clear explanations for, the curious phenomenon of 'explosion limits', confining explosive reaction, at any fixed temperature, sharply to a particular pressure region. Elaborate studies of the effects of nitrogen peroxide and other foreign gases on the hydrogen-oxygen reaction brought to light the very great kinetic complexities of an apparently simple type of chemical change. In this work Hinshelwood drew attention to the influence of the container surfaces on chain reactions, and also clarified the confusion of evidence concerning the effects on reaction rates of the intensive drying of gases. He discovered the inhibition of certain gaseous reactions by nitric oxide and interpreted the effect as due to the removal of radicals from, and the suppression of, 'chains'.

Hinshelwood has also carried out a large number of experiments on heterogeneous reactions and shown that their differences in kinetic behaviour can be explained by the application of the concepts of Langmuir.

Throughout all these researches, carried out with the utmost economy and directness, though with full experimental precautions, and interpreted in the most lucid manner, Hinshelwood has never lost sight of the essential complexity of chemical reaction mechanisms. He has always been ready to modify his views in accordance with new experimental evidence, and to make the fullest use of the more recent developments of wave-mechanics and of statistical mechanics. Summarized by their author in two well-known treatises, Professor Hinshelwood's distinguished researches furnish abundant ground for the award to him of the Davy Medal.

The DARWIN MEDAL is awarded to Professor D. M. S. WATSON, pre-eminent among palaeontologists for his contributions to knowledge of the course of vertebrate evolution. His researches have been concerned mostly with the origin of the land vertebrates, with the fishes most nearly related to them, and with the main line of evolution leading to the mammals.

It will not be possible on this occasion to survey Watson's work in all its aspects, and mention must be restricted to some of the major lines of advance which it has opened. His Croonian Lecture, in 1925, summarized the conclusions which he had reached by that date as to the evolution of the Amphibia, demonstrating for the first time the relationship of the Stegocephalia to the Osteolepid fishes.

In addition to tracing the descent of land vertebrates thus from Amphibia back to fishes, Watson followed the line of the evolution of the mammals, through early, primitive reptiles, the Cotylosaurs, to the mammal-like reptiles, in a large series of valuable papers. He related this work on the reptiles to that on the amphibians in a paper on the evolution of the shoulder girdle of the Tetrapoda.

In this work on the fossil vertebrates, in its relation to the course of evolution,

Watson has not confined his attention to morphological details, but, with an enterprise remarkable in a palaeontologist though characteristic of his outlook, has considered where possible the functional significance of the structures preserved in the rocks; as in the paper in which he considers the mode of action of the shoulder girdle and deduces the nature of the musculature of a group of marine fossil reptiles.

Pursuing his study of mammalian origins, Watson was led to study the most primitive of living mammals, the oviparous Monotremes, and to discover that characters in which their skulls differ from those of other mammals can be regarded as extreme developments of features observed in the skulls of certain fossil, mammal-like reptiles.

Watson's work has continued in full vigour into recent years, and has brought two further contributions of major importance to the study of evolution in the vertebrates. One is concerned with the origin of the frogs from more primitive amphibian types, while the other shows that a group of fishes from the Old Red Sandstone constitute a separate class of vertebrates, equal in rank to and ancestral to the remaining fishes.

Tracing, in this brilliant series of researches, the main stages of the descent of the mammals from their earliest fish-like ancestry, Professor Watson has certainly performed 'work of acknowledged distinction in the field in which Charles Darwin himself laboured'.

The BUCHANAN MEDAL is awarded to Sir WILSON JAMESON, formerly Dean of the London School of Hygiene and Tropical Medicine and since 1940 the Chief Medical Officer to the Ministry of Health and the Board of Education. In both capacities Jameson has shown himself to be a man of stimulating influence and leadership, determined and persistent in his efforts to ensure that advances of medical knowledge in the laboratory, the clinic and the field shall receive prompt application in administrative practice.

Largely to Jameson's vigorous policy is due the hope that active immunization against diphtheria, which has banished the disease from many large communities of North America, will at length find systematic and effective application in this country, where many of the discoveries were made which have rendered it safe and practicable. In the prompt official adoption of methods using modern technical resources, to deal with the recent increase of tuberculosis under war conditions, and in the recognition of adequate and scientifically planned nutrition of the people as a central item of an effective health policy, Jameson's active and enlightened influence can again be discerned.

Of the grounds on which the founders of the Buchanan Medal desired the awards of it to be made, Sir Wilson Jameson's high claim to it is based on 'administrative and constructive work' of outstanding merit in the service of Hygienic Science.

The HUGHES MEDAL is awarded to Professor ENRICO FERMI, now of New York. Professor Fermi has made most notable contributions both to theoretical and

experimental physics. In the early days of the modern quantum theory he was one of the first theoretical physicists to appreciate the generality of the considerations put forward by Pauli and known as the Exclusion Principle. This led him to discuss the statistical theory of a perfect gas of particles in equilibrium, obeying this principle, with results which were obtained independently and almost simultaneously by Dirac by similar methods. These results of Fermi and Dirac are of the utmost importance in the modern theory of assemblies of similar particles, such as electrons, protons, and neutrons. Following this outstanding personal contribution, Fermi played a great part in building up at Rome a distinguished school of theoretical physics, where he himself made one of the earliest successful attempts to construct a theory of radioactive β -ray change. This theory shows the most profound insight into the theoretical nature of the quantum theory.

His interest in the atomic nucleus led Fermi naturally on to his experimental studies in this field. Immediately after the discovery of the neutron he realized that it provided a new possibility of attack on the nucleus and of stimulating nuclear change by neutron bombardment. This work opened up the fruitful modern field of study concerned with the transformations of nuclei of medium and great atomic number, and led directly to the most exciting transformations of all, the nuclear fission of uranium and thorium.

Professor Fermi's work is characterized throughout by profound insight and great experimental skill. In the fields which he has made his own he is universally acclaimed a leader.



ISAAC NEWTON

from a portrait by Charles Jervas, in the possession of the Royal Society. This was presented by Sir Isaac Newton at the meeting held on 16 May 1717, and is recorded in the Journal Book under that date in the words: 'The President gave the Society his picture drawn by Mr Jervase for which he had their thanks'.

**NEWTON
TERCENTENARY
CELEBRATIONS**

**AT THE ANNIVERSARY MEETING
OF THE ROYAL SOCIETY
30 NOVEMBER 1942**

**PRESIDENTIAL ADDRESS: BY SIR HENRY DALE
NEWTON AND THE SCIENCE OF HIS AGE: BY E. N. DA C. ANDRADE
NEWTON AS AN EXPERIMENTER: BY THE LORD RAYLEIGH
NEWTON AND THE SCIENCE OF TODAY: BY SIR JAMES JEANS**

EXHIBITS

ACKNOWLEDGMENTS

Address of the President
Sir Henry Dale, C.B.E.
Anniversary Meeting, 30 November 1942

We are to-day within a few weeks of the three hundredth anniversary of the birth of Isaac Newton. Wherever the progress of our Western science and philosophy has become effective, men will remember what that event was to mean for the world. Newton, as we shall hear, at the age of 43, when he had determined to abandon all further concern with natural philosophy, was induced at length, by Halley's friendly insistence, to give written form and system to the mathematical discoveries with which his amazing mind had been occupied over a period of some twenty years. The result was one of the greatest intellectual achievements in the history of mankind—the *Principia*, providing for more than two centuries a framework for the mechanical interpretation of the universe and a basis for the building of physical science, and therewith of the material structure of our modern civilization.

We in Britain regard Isaac Newton as still, beyond challenge, the greatest of our men of science. Nor should the claim be limited to this island or to the British Commonwealth of Nations; for it was not till nearly half a century after Newton's death that former British colonists in North America began their development of an independent nation; and Newton is theirs as well as ours.

But, while we may proudly claim him as the countryman of all who share the birthright of the English tongue, the discoveries of science have belonged, and must belong again, to the whole world, and Newton's achievement is a part of the common heritage of all peoples. It cannot be doubted that, if it had fallen in normal times, this tercentenary would have been marked by the greatest of international gatherings, in which men of science and philosophers from all the world would have assembled to do honour to Newton's memory. It would have been natural then to expect leadership, in such an enterprise, from the only two institutions which were intimately concerned with Newton's career as a man of science—Trinity College, in Cambridge, and this our Royal Society of London. Our two foundations did, indeed, confer as to the wisdom of attempting by joint action, even in this year of war, to arrange such a restricted and domestic celebration as the present conditions would allow. We agreed, however, to put aside such planning for the present, carrying it forward in our hopes to the time when a world at peace may be able to join in international commemoration of an event which has meant so much for all mankind. May the time be not too far distant.

To-day we are holding the 280th Anniversary Meeting of the Royal Society, on Saint Andrew's Day as by regular custom, ever since the first meeting on that day in 1662. It seemed to us that we should depart, on this occasion, a little from the

usual order of our proceedings, so that, on a day so near to the tercentenary of his birth, our Fellows and our guests may be reminded of Newton. We have accordingly asked three of our Fellows to address the Society on different aspects of Newton's work, in its relations to the science of the past and the present. We have asked Professor Andrade to give us the opportunity of understanding the magnitude of the change which Newton's work produced, in the conceptions of the material universe which were current in his own times. We know that to his contemporaries Newton's discoveries came as a great revelation, and we hope that Professor Andrade will help us to understand why they did so. We have asked Lord Rayleigh to deal, by demonstration where possible, with the experimental work of Newton and the great discoveries which he made by that method. This is an aspect of his greatness which popular estimates have tended to overlook; but I think that Lord Rayleigh will be able to convince us that Newton as an experimenter would have had claims to a place among our greatest men of science, even if he had failed, as he so nearly did fail, to write the *Principia*. And, finally, we have asked Sir James Jeans to give us some reassessment of the validity and permanence of Newton's system, in relation to the immense advances of knowledge in our own times. There are many who have not the mathematical equipment to follow them in detail, who are nevertheless aware that revolutionary changes have been taking place in conceptions of the mechanics of the universe and of its ultimate material units. How is the Newtonian system affected by the discoveries which have required the general theory of relativity and the quantum mechanics at opposite ends of the stupendous scale? Is it being supplemented, modified or superseded after its centuries of dominance? We hope that Sir James Jeans will tell us; and we may remember, perhaps with comfort, that Newton's *Principia* seemed difficult and abstruse to his contemporaries, and that he even confessed that he had made it so deliberately, 'to avoid being bated by little smatterers in mathematics'.

Before I call on our chosen lecturers to address us, there are two other matters relating to Newton and the Royal Society, which it seems proper to mention here. In the hamlet of Woolsthorpe, near Colsterworth on the Great North Road, some six miles south of Grantham, there is still a modest manor farm-house, with a small orchard in front of it. Here the Newtons lived, simple yeoman farmers, and here, two months after his father had died, Isaac Newton was born, a puny, premature infant, on Christmas day 1642, twenty years before the Royal Society was incorporated by the grant of its first Charter. The house stands but little altered since that day. The room in which Newton was born has a simple marble tablet on the wall, inscribed with Pope's well-known couplet. But this house had importance in Newton's later life and in his work, and not only as his birthplace. It was here that he returned from his schooling at Grantham, at the age of 18, to take charge of the farm for his mother; and here, to the incalculable gain of science and the world, he showed such incompetence as a farmer that he was sent back to school, and thence to Cambridge. It was here, again, that he returned in the autumn of 1665, when the plague drove him from Cambridge; and here, during the following

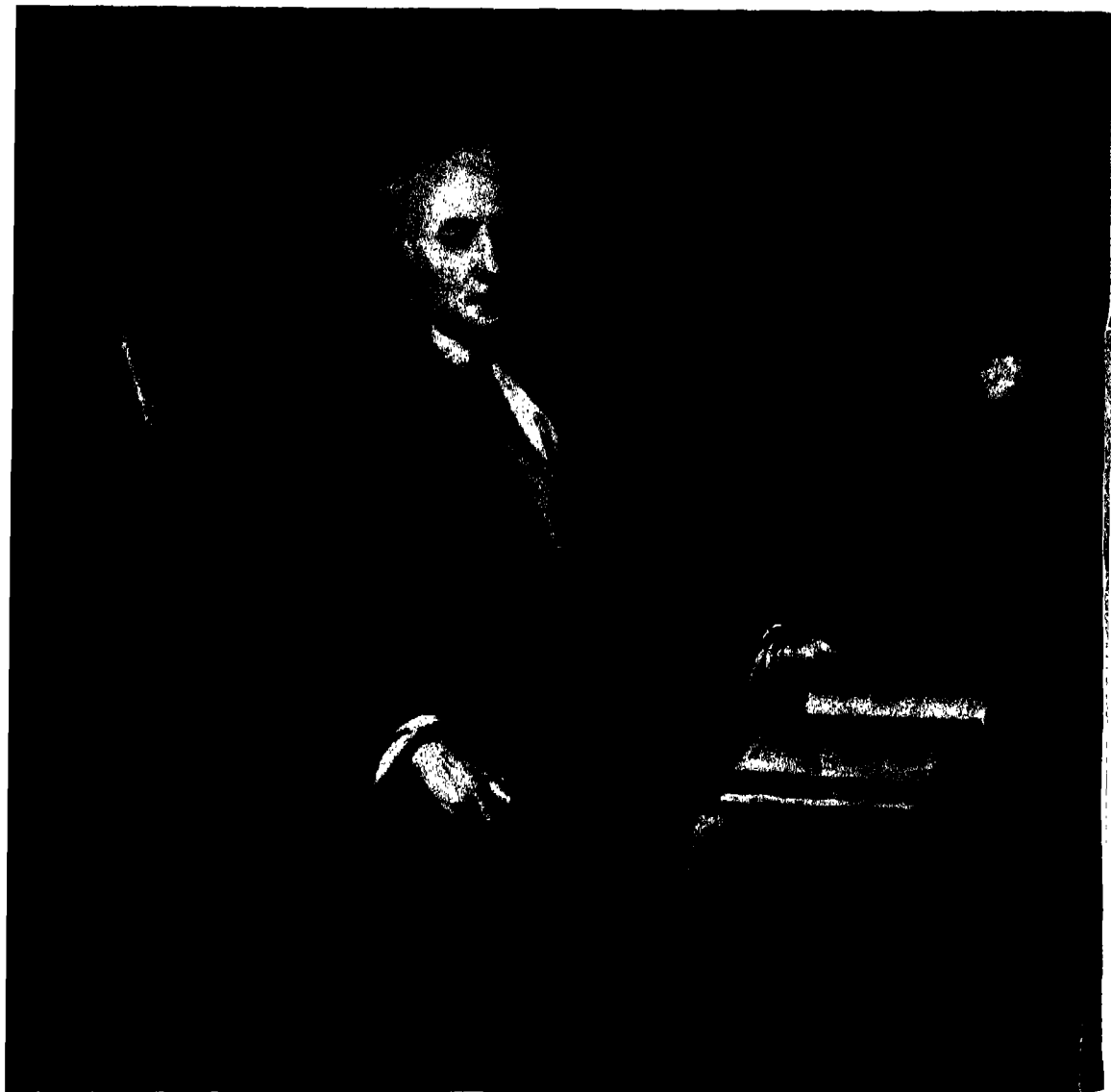
eighteen months of quiet exile in the country, his early ripening genius grasped already the essential principles of his major theoretical discoveries. You can still see the upper chamber which he then used as a study; and in the little orchard there is an old, recumbent apple tree, which, they will tell you, is descended by direct grafting from that which Newton saw. The land which Newton's family farmed is rapidly being laid waste, alas, by quarrying for ironstone; and soon there will be little left unspoiled save the orchard and garden round the house. It has seemed to us that, in this year of commemoration, something should be done to preserve for posterity a house and garden which carry such momentous memories, and which have meant so much for science. We have accordingly formed a small Committee, in which Sir John Russell and Sir James Jeans have joined with the Officers of the Royal Society. We have been in friendly negotiation with the lord of the manor, Major Turnor—a name associated in several past generations with the Royal Society and with Isaac Newton—as to the possibility of acquiring this now tiny but historic property, so that it may be put for as long as possible beyond the risk of damage or decay. Major Turnor has generously offered to sell the property for this purpose at a price substantially less than its value, and only this morning I received a letter from Lord Macmillan which enables me to announce that the Pilgrim Trust will be responsible for the sum required for the purchase.

Then I think that it is our special duty here, at this Anniversary Meeting, to remember that, while Newton's great discoveries belong to the world, they came to publication through the Royal Society, and that Newton occupied its presidential Chair for the last twenty-four years of his long life. Though his *Opticks* was not published till after he had become President, his original work for science was practically finished by the time of his election, and he had for some years been Master of the Mint. There can be no doubt, however, that the wide fame of his achievements, and the respect and admiration in which he was everywhere held, did much, at a critical period in its history, to establish the prestige of this Society in the eyes of the world. Let us then remember to-day that Isaac Newton, the greatest man of science of our race, was also the greatest of the Royal Society's presidents.



WOOLSTHORPE MANOR

from photographs taken by Walter Lee of Grantham, November 1942.



ISAAC NEWTON

from a portrait by Johan Vanderbank in the National Portrait Gallery.

Newton and the science of his age

BY PROFESSOR E. N. DA C. ANDRADE, F.R.S.

It is my task—my honourable and inspiring task—to say something of Isaac Newton as seen against the background of the science of his time. I shall try to display briefly the position as he found it and to resume in a small space his great achievements and the changes in outlook which they produced. In praising Newton I shall endeavour not to do injustice to his great forerunners and to the men of his time who pursued worthily the same great ends as he did, and who would have held the centre of the stage in any other age than that dominated by him. For Newton, like Shakespeare, did not stand as a lonely adventurer into new realms, though he travelled further and straighter than the rest. Shakespeare was the supreme poet and playwright at a time when poetry and plays were part of the life of every cultivated man and occupied the attention of the brightest intellects. Newton was the supreme scientist in an age when the quantitative method of questioning Nature was abroad in the air. Each was the child of his time.

Let us consider the position when Newton went to Cambridge in 1661. The hold of Aristotle, whose works had for centuries been the ultimate resort of all those seeking knowledge of the working of nature, had been shaken off by such men as Galileo and Gilbert, but most of the learned still thought that those who relied on experiment were pursuing a futile and impudent course. The first resolve of Marlowe's Faustus

Having commenc'd, be a divine in shew,
Yet level at the end of every art,
And live and die in Aristotle's works

still represented the aim of many students. The foundation of the Royal Society in 1662 had been the occasion of many attacks on the experimental method, attacks stoutly met by Glanvill and by Sprat, and as late as 1692 Sir William Temple's *Essay upon the Ancient and Modern Learning*, satired by Swift in *The Battle of the Books*, set out to prove the superiority of the philosophers of the ancient world over all the moderns. Thus when Newton was a young man the new experimental method of questioning Nature was steadily making its way and the omniscience of the ancients was being called in doubt by a new school, but experimental science was by no means firmly established as a respectable study.

The great figures among the worthies of the exact sciences who had already appeared at that time were Copernicus, Tycho Brahe, Kepler, Gilbert, Galileo and Descartes. Kepler, following his great forerunners, had found the true laws of planetary motion, which were to be explained by Newton. Kepler's views as to the mechanism of the planetary motions were in his earlier writings largely mystical,

involving the perfect properties of the five regular solids and also certain motivating souls or spirits, *animae motrices* (figure 1). Throughout he held to the medieval point of view that a body could not maintain its motion unless there were a force propelling it. In his later writing he invoked a magnetic force, but

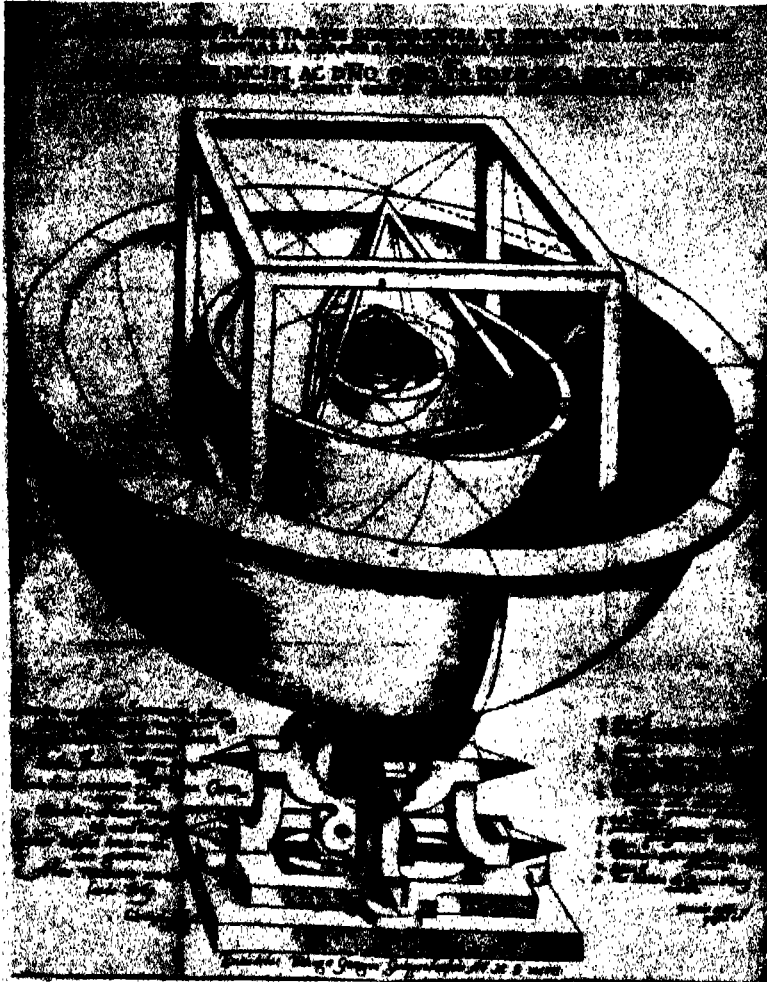


FIGURE 1. Illustration from Kepler's *Mysterium Cosmographicum*, 1596, showing the orbits of the planets fitted between the octahedron, icosahedron, dodecahedron, tetrahedron and cube.

it was not directed to the sun, like the true gravitational force, but pushed the planets on their way—*non est attractoria sed promotoria*. It was essentially bound up with the rotation of the sun. Thus he was ignorant of the basic laws of mechanics and his magnetic force had none of the true properties of magnetic forces. Kepler made no approach to a mechanical explanation of his laws.

Gilbert had not only established the basic principles of terrestrial magnetism and carried out fundamental work on electricity, but had invoked a force from the moon—a magnetic force it is true—to produce the tides. Galileo's greatest achievement had been to lay the foundations of mechanics. None of these men, however, had made any impression on the bulk of the learned: Francis Bacon, for instance, neglects Galileo and Kepler, and refuses to take Gilbert seriously. The great figure in the eye of natural philosophers was Descartes, who had developed a cosmogony based upon mechanical principles, not precise mechanical principles it is true, but principles very different from the mystical ones then in vogue. Joseph Glanvill used to lament that his friends had not sent him to Cambridge, where he might have learned the new philosophy of Descartes, rather than to Oxford, where Aristotelianism ruled. Descartes' system had acquired such a hold on men's minds that his views were still supported long after Newton's death. Both Jean Bernoulli, who died in 1748, and Fontenelle, who died in 1757, were Cartesians to the end. The demolition of the Cartesian system was, in contemporary eyes, one of Newton's greatest achievements. James Thomson said in his *Ode to the Memory of Sir Isaac Newton*, published immediately after his death:

The heavens are all his own; from the wild rule
Of whirling vortices, and circling spheres,
To their first great simplicity restored.
The schools astonished stood.

From the *Principia* it is clear that Newton himself derived particular satisfaction from having invalidated the Cartesian system. It is fitting, then, that we start our consideration of contemporary science by a glance at this system, not only because it was the only attempt before Newton to explain the motions of heavenly bodies on general principles, but because it furnishes a contrast which brings out the essential Newtonian point of view.

Descartes starts, in the spirit of medieval thinkers, from certain general philosophical principles. He decides that the fundamental property of matter is extension—impenetrability, colour, hardness and so on are only secondary characteristics. Extension, which has three directions, is the subject of mathematics: motion is the subject of mechanics. All the different qualities of different kinds of matter are provided by different motions of the minute parts of which it is composed. 'Give me extension and motion' declares Descartes, 'and I will construct the world.' One consequence of his fundamental belief is that there cannot be a vacuum, for extension without matter is a contradiction. He further considers, on theological grounds, that the quantity of motion in the heavens must be constant. He blames Galileo for founding his mechanics on experiments and not on reflexions on first causes. 'Everything Galileo says about the philosophy of bodies falling in empty space is built without foundation: he ought first to have determined the nature of weight.' Newton's point of view, of course, was the exact opposite to that of Descartes: he says in the famous letters to Bentley, '...for the cause of gravity is what I do not pretend to know, and therefore

would take some time to consider of it', and again, 'gravity must be caused by an agent acting constantly according to certain laws, but whether this agent be material or immaterial, I have left to the consideration of my readers'. For Newton, as for the best of his successors, science was concerned with the question of 'How?': Descartes, like the ancients, was concerned with the insoluble question of a fundamental 'Why?'.

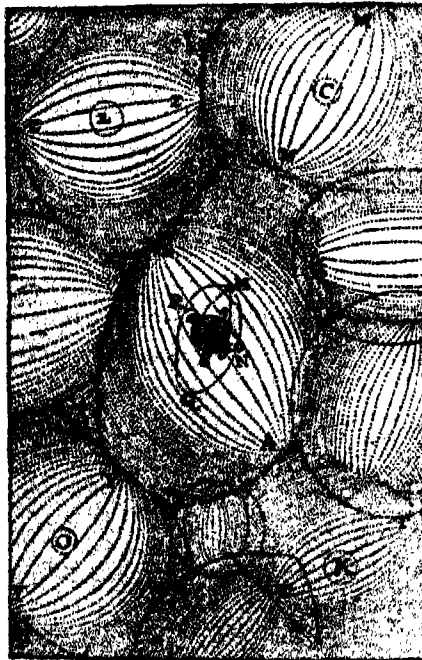


FIGURE 2. A page from Descartes' *Principia philosophiae*, 1644, showing in the middle the sun's vortex, governing the planets. Other vortices surround it.

It followed from the philosophic hypothesis of Descartes that the only kind of motion possible in a plenum was a motion in closed paths, more particularly a circular motion, since a particle could only move if another particle took its place. It was on grounds of this kind that he elaborated his vortex hypothesis. Certain very fine particles, which filled interplanetary space, moved round ceaselessly in huge vortices and carried the planets with them. The moon was carried round the earth by a minor vortex, and so on (figure 2).

Descartes' cosmogony, then, was founded on a philosophical system: it was pictorial and unquantitative. The paths of the comets, handed on from one vortex to another, were in particular irreconcilable with observation. There was no attempt to deduce Kepler's laws, or to show how anything but circular motion could result from the vortices. The whole scheme was spun from the brain of Descartes, with more or less casual references to actual phenomena. No doubt it was referring to Descartes that Roger Cotes said in his Preface to the second edition of the *Principia*, 'Those who fetch from hypothesis the foundation on which

they build their speculations may form indeed an ingenious romance, but a romance it will still be.' In spite of, or perhaps because of, this, Descartes' influence was immense, and when Newton was at Cambridge as a young man it was Descartes who was the great authority for all such progressive spirits as speculated on the structure of the universe.

We now turn to the astonishing story of the birth of the *Principia*. The time was ripe for the appearance of this great work. As regards the laws of motion, Galileo, whose services Newton freely acknowledges, and Descartes himself had done much to prepare the way for the more precise and particular formulation which Newton gives. Hooke among others had clearly expressed the protest of the most forward spirits of the time against the speculative method. 'The truth is, the Science of Nature has been already too long made only a work of the Brain and the Fancy: it is now high time that it should return to the plainness and soundness of *Observations on material and obvious things*.' The Royal Society was active in stimulating the pursuit of the new method and on the Continent the *Accademia del Cimento* had done excellent work, while the *Académie des Sciences* was founded in 1666. These continental societies had little influence on Newton, but are symptomatic of the general movement towards the experimental method. The stage was set for great things.

The story opens at Woolsthorpe, Newton's birthplace, whither Newton had come from Cambridge in June 1665 to escape the plague. He was then 22 years old and not yet a Master of Arts or Fellow of Trinity. He had read what he calls 'Schooten's *Miscellanies*' (probably the *Exercitationum Mathematicarum Libri V*), Descartes' *Geometria* and Wallis' works, and further was, of course, familiar with the work of his teacher Barrow. He had written his first treatise on the calculus, or 'fluxions' as he called it, but he had published nothing. The words which he wrote some fifty years later about this great springtime of his intellectual life have often been quoted but cannot well be omitted on an occasion like this. 'And the same year (1666) I began to think of gravity extending to the orb of the Moon, and having found out how to estimate the force with which a globe revolving within a sphere presses the surface of the sphere, from Kepler's Rule of the periodical times of the Planets being in a sesquialterate proportion of their distances from the centers of their orbs I deduced that the forces which keep the Planets in their Orbs must [be] reciprocally as the squares of their distances from the centers about which they revolve: and thereby compared the force requisite to keep the Moon in her orb with the force of gravity at the surface of the earth, and found them answer pretty nearly. All this was in the two plague years of 1665 and 1666, for in those days I was in the prime of my age for invention, and minded Mathematicks and Philosophy more than at any time since. What Mr Hagens has published since about centrifugal forces I suppose he had before me.'

It seems likely that he had already had his laws of motion in his head—in any case it is clear that he was convinced that every body would continue to move uniformly in a straight line unless some force acted on it, and that, therefore,

there must be some force acting on the moon which drew it away from the straight line, tangential at any moment to its path, in the direction of the earth. Treating the moon's path as circular, from Kepler's third law, and from the law connecting the centrifugal force with the radius and the velocity, or the equivalent proposition to which Newton refers, it is easy to deduce the inverse square law. To show that the force keeping the moon in her orbit is the earth's gravitational force, assumed to diminish as the inverse square, is, however, a further step demanding a computation of how strong the gravitational force at the moon's orbit will be, compared to the measured force at the surface of the earth. Newton made this step and found it 'answer pretty nearly'. Why, then, did he delay the announcement of the law of gravity for twenty years or so?

There is always a ready answer to questions of this kind where Newton is concerned—that he never published anything until invited, in general strongly urged, to do so. It is, however, clear from many signs that Newton was not himself satisfied about the matter until some time about 1686. The usual story is that he took a wrong radius for the earth, namely one corresponding to 60 miles for 1° of latitude instead of the correct value of about 70, but this story is very improbable on many grounds, one of which is that good values were readily available to him.

The real reason for Newton putting the work aside seems to have been that the calculation, as far as the force at the earth's surface is concerned, depends essentially upon it being legitimate to assume that the earth's mass may be considered as concentrated at the centre. That this assumption is valid is far from obvious. It is fairly clear from certain passages in *De Motu* and in the *Principia* that it gave Newton some trouble to prove this assumption and that he did it late. In a letter to Halley of 20 June 1686 he says, 'I never extended the duplicate proportion lower than to the superficies of the earth, and before a certain demonstration I found the last year, have suspected that it did not reach accurately enough down so low.' Although this refers to the gravitational force within a sphere, this and the point under discussion are involved in the same mathematical demonstration, which Newton gives in the *Principia* in Book I, proposition LXXI and other propositions following it. In any case Newton appears not to have been satisfied with his first calculations and to have turned to other things, possibly his optical experiments.

It is a strange thing that the *Principia* owes its publication largely to a quarrel with Hooke, and its sequel. In 1679 Hooke, then acting as Secretary of the Royal Society, wrote to Newton about various scientific matters and asked him very civilly for a philosophic communication—a paper, as we should say nowadays. He also asked for Newton's opinion on his *Potentia Restitutiva* and 'particularly if you will let me know your thoughts of that [hypothesis] of compounding the celestial motions of the planets of a direct motion by the tangent and an attractive motion towards the central body'. In his reply Newton made an extraordinary remark. 'But yet my affection to philosophy being worn out, so that I am almost as little concerned about it as one tradesman uses to be about another man's trade or

a countryman about learning, I must acknowledge myself averse from spending that time in writing about it which I think I can spend otherwise, more to my own content and the good of others: and I hope neither you nor any body else will blame for this averseness.' This is but one of many occasions on which Newton expresses his disinclination, almost distaste, for any further scientific work, his first antipathy having been aroused by the disputes and misunderstandings consequent on the publication of his first great paper on the prism.

To return, Newton did comply with Hooke's request for something for the Society by pointing out that a body let fall from on high should strike the earth slightly east of the perpendicular, and gave precise and excellent directions for carrying out the experiment. The purpose was to prove the diurnal rotation of the earth. Hooke, in reply, pointed out that the ball should fall to the south as well as to the east, and further corrected Newton in a point which is too complicated to discuss here and one on which differences of statement can be due to different interpretations of the problem. This correction, tactlessly expressed, irritated Newton in the highest degree and he answered curtly. In further letters, written in apparent unconsciousness of the annoyance he had given, Hooke suggested that the law needed to explain the planetary motions was the inverse square law.

Now other men had come to the same conclusion. In particular, Wren and Halley had discussed with Hooke the possibility of explaining the mechanism of the heavens on the basis of an inverse square law. Hooke declared that he could demonstrate mathematically that the path of a particle in a central inverse square field would be an ellipse, but it is clear that he never did so, and equally clear that he had not the mathematical equipment necessary to begin an attack on the problem. He was in the unfortunate position of being entirely convinced of a truth that he could not prove. In August 1684 Halley visited Cambridge and asked Newton what the path would be. He replied that it would be an ellipse and that he had formerly calculated it. He could not find the calculation but soon sent a proof (or apparently two different proofs) to Halley.

This incident seems to have aroused Newton from the distaste for science into which he had fallen, and he put together the treatise *De Motu*, founded on a course of lectures, which Halley presented to the Royal Society on 10 December 1684. The story of how Halley then coaxed and cajoled Newton into writing the *Principia* is familiar, but we in this Society ought not ever to celebrate the great work without a tribute to Halley, who not only realized at once the fundamental importance and significance of Newton's work but used all his tact to get the book written and made himself financially responsible for the production, the Royal Society being in financial difficulties at that time. We are not now in financial straits and shall, I believe, be glad to bear the expense of producing a second *Principia* when the genius of our age brings it forth. The book appeared in 1687, and bears the imprimatur of the then President (figure 3), who, although he achieved nothing in science, is still remembered. He was Samuel Pepys.

The *Principia* is not an easy book to read. The proofs are all given in form of classical geometry, although, since it is certain that at the time when it was written Newton was in possession of the fundamental processes of the calculus and of the methods of analytical geometry, it is unlikely that this was the form in which he first derived them. Figure 4 shows a typical geometrical diagram, dealing with the moon's motion, and a comparison with the pictures from Kepler and Descartes gives a graphic representation of the changed spirit in handling the mechanism of the heavens. With reference to the mathematical methods of the

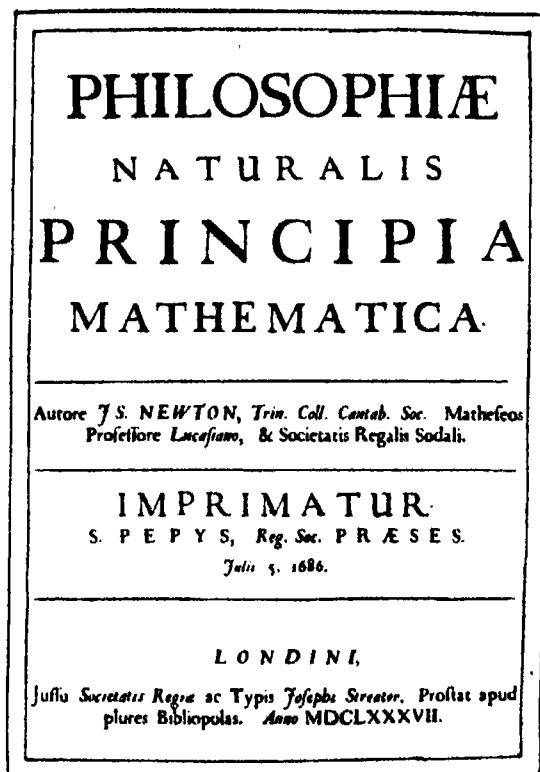


FIGURE 3. Reduced facsimile of the title page of the *Principia*, bearing the imprimatur of Samuel Pepys.

Principia Whewell has said 'Nobody since Newton has been able to use geometrical methods to the same extent for the like purposes; and as we read the *Principia* we feel as when we are in an ancient armoury where the weapons are of gigantic size; and as we look at them we marvel what manner of man he was who could use as a weapon what we can scarcely lift as a burden.' Various conjectures have been made by Rosenberger, Cantor, Giesel, Gerhardt and others as to why he did not use the new methods: Rouse Ball thinks that he was probably unwilling to add to the difficulties by introducing a new mathematical method. Be that as it may, he had a horror of unfounded criticism, and, as he told Dr Derham 'to avoid being baited by little smatterers in mathematics,

A complex geometric diagram featuring a sphere with several intersecting great circles. Key points are labeled with letters: H , K , P , Q , N , G , J , T , A , I , Z , n , q , a , m , l . Lines connect these points, some passing through the sphere and others outside. Symbols S and O are placed near the left edge of the sphere. The diagram illustrates a geometric proof or construction, likely related to astronomy or optics.

FIGURE 4. A typical geometrical diagram from the *Principia*.

In the first book the motions are all supposed to take place in a non-resistant medium. In the second Newton considers motions in a resisting medium, always with his eye on the Cartesian world system which he was intent to demolish. He deals with a resistance proportional to the velocity and with a resistance proportional

* Andrew Motte's translation of 1729, volume 2, page 201.

to the square of the velocity, and further points out different kinds of fluid resistance, which he later defines most clearly in the 28th Query appended to the third edition of the *Opticks*, 'for the resisting Power of the fluid Medium arises partly from the Attrition of the Parts of the Medium and partly from the *Vis inertiae* of the Matter', that is, partly from the viscosity and partly from the bulk motion of the medium. The latter resistance he assumes proportional to the square of the velocity. In this book he opens the way to the hydrodynamics of real fluids. In the part dealing with hydrostatics he proves the law of the diminution of atmospheric pressure with height. He then further discusses the motion of the pendulum and is the first to suggest its use for making a survey of the gravitational acceleration. In another connexion he derives an expression for the velocity of sound, this being the first case of the calculation of the velocity of a wave from the properties of the medium. The only other point in this book to which I will refer is the calculation which Newton carries out on the motion of an infinitely extended viscous fluid in which a body rotating upon its axis is immersed. What he has in mind is the Cartesian vortex—'I have endeavoured in this proposition to investigate the properties of vortices, that I may find whether a celestial phenomenon can be explained by them.' He effectively defines the force 'arising from the want of lubricity in the parts of the fluid' as proportional to the velocity gradient, whence the term 'Newtonian viscosity'. He is thus the first to touch the mathematics of viscous fluids. He finds that the periodic time of circulation of the fluid carried round by a rotating sphere is proportional to the square of the distance from the centre of the sphere,* which is grossly inconsistent with Kepler's third law. Newton considers that his deduction is a clear refutation of the Cartesian vortices and, for once, almost gloats over his victory. 'Let philosophers then see how the phenomenon of the sesquiplicate ratio can be accounted for by vortices.' He brings many other objections against the Cartesian vortices: he points out, for instance, that a continuous supply of energy will have to be given to the sphere to maintain the motion, because 'it is plain that the motion is proportionally transferred from the centre to the circumference of the vortex, till it is quite swallowed up and lost in the boundless extent of that circumference'.

The second book of the *Principia* is the foundation stone of mathematical physics. In it Newton shows an extraordinary instinct for grasping the essentials of a problem—it would almost seem that he knew the solution in advance and added the proof as a concession to those less clear-sighted. As William Whiston, who knew him well, and succeeded him in the Lucasian chair, says: 'Sir Isaac, in mathematics, could sometimes see almost by intuition, even without demonstration....And when he did but propose conjectures in natural philosophy, he almost always knew them to be true at the same time.' The manner of his refutation of the Cartesian vortices introduces a new spirit into the discussion: no hypothesis

* Actually this is an error, which appears never to have been pointed out, although Stokes has indicated the like error in the case of the rotating cylinder, which Newton works out. It should be the *cube* of the distance.

about the heavens is tenable unless the quantitative deductions from it agree with observation.

The third book opens with an introduction where, after stating what has been done in the first two books, Newton sets down the superb sentence—'Superstat ut ex iisdem principiis doceamus constitutionem systematis mundani'—it remains that from the same principles we demonstrate the form of the system of the world. It is this third book which based celestial mechanics so firmly that what was done in the next two hundred years was rather extension of, and improvements on, the Newtonian method than anything radically new. Not only does Newton establish the movements of the satellites of Jupiter, Saturn and the Earth, and of the planets round the Sun (or rather, as he points out, round the centre of gravity of the solar system) in terms of his gravitational theory, but he shows how to find the masses of the sun and planets in terms of the earth's mass, which he estimates quite closely; he accounts for the flattened shape of the earth and other planets; calculates the general variations of g over the surface of the earth; explains the precession of the equinoxes by consideration of the non-sphericity of the earth; calculates the main irregularities of the motion of the moon and of other satellites from the perturbing effect of the sun; explains the general features of the tides; and finally treats the orbits of comets in a way that shows that they are members of the solar system and enables the return of Halley's comet in 1759 to be accurately calculated. This brief and imperfect catalogue is merely a reminder of the scope of this extraordinary book, which drew from Laplace, no enthusiast, '... all this, presented with much elegance, assures to the *Principia* preeminence over all the other productions of the human mind'. The book closes with the famous General Scholium which returns to the confutation of Descartes' vortices and says, concerning the cause of gravity, 'Hypotheses non fingo'.

We have a portrait of Newton, by Kneller, at about the time of the publication of the *Principia* which is particularly impressive. In contrast to most of the later and more formal portraits we see him in his own hair and in the casual clothing which we may suppose him to have worn when at work. The look of wild, almost hostile remoteness and of dominating and piercing intelligence seem to show that the artist has well read the features of his sitter and given us a true picture of the man in his hours of creative thought.

Although the book was eagerly bought, the Newtonian method and discoveries made way but slowly. Biot says that of Newton's contemporaries three or four only were capable of understanding the *Principia*, that Huygens only half adopted the ideas, Leibniz and Jean Bernoulli fought against them, and that fifty years had to pass before the great truth demonstrated by Newton was understood by the generality of men of science, let alone developed. No doubt the difficulty of the book had much to do with the tardy appreciation, outside a narrow circle, of its contents. The Cartesian scheme was easy, pictorial, general: the Newtonian difficult, mathematical, precise. The very method of attacking the problem was altogether new. The second edition appeared in 1713, edited by Roger Cotes, of

whom Newton said 'if Mr Cotes had lived we might have known something'. The preface clearly shows that among the learned the Newtonian scheme had not been widely accepted. On the Continent the progress was still slower and it is generally held that it was Voltaire's *Elémens de la Philosophie de Newton*, which appeared in 1738, that led to Newton's work being appreciated in France. Incidentally, the story of the apple, which Voltaire had from Newton's niece, Mrs Conduitt, appeared in the second edition of Voltaire's book, in 1741. Later, it was in France that Newton's work was raised to great glory, when Lagrange and Laplace erected edifices of splendour and elegance on Newtonian foundations.

The optical work of Newton was perhaps his favourite study: at any rate it was that to which he made what is probably the only enthusiastic reference which he ever permitted himself, when he said, of his fundamental discovery in the matter of prismatic colours, 'being in my judgment the oddest, if not the most considerable detection which hath hitherto been made in the operations of nature'. The beautiful experiments which formed the background of this work will be dealt with by Lord Rayleigh, so I shall say nothing of them, but a few words as to the Newtonian attitude and theory may be permitted.

The nature of colour had from the days of Aristotle been the subject of philosophical speculation. According to Aristotle, colours are a mixture of light and darkness, or of white and black, a view which, embellished and modified in various ways, survived Newton and appeared again in, for instance, Goethe's writings. Descartes was apparently the first to break really new ground in comparing colours to notes in music: his view of light was that it was a pressure transmitted through the particles that filled all space, and he conjectured that a rotation of the particles might be the effective cause of colour, a view which presented inherent difficulties effectively exposed by Hooke. We may agree with Huygens that 'Descartes has said nothing that is not full of difficulties or even inconceivable, in dealing with light and its properties'. The medieval question as to whether light was a substance or an accident was still occupying the attention of even acute experimentalists such as Grimaldi, the discoverer of diffraction, whose book appeared in 1665. It was Hooke again, who, with his extraordinary flair for the truth, combined with an inability to overcome the last difficulties that stood between him and a convincing conclusion, initiated the modern views. In his *Micrographia*, 1665, he expressed the view that light was a very quick vibration 'propagated with a finite velocity. He gave what resembles Huygens' construction for finding the wave front on refraction, but, since he thought that light travelled faster in a solid medium, e.g. glass, than in air, he found that in the medium the wave front must make an acute angle with the ray. It was with this 'obliquity' that he connected colour. 'Blue is an impression on the Retina of an oblique and confus'd pulse of light, whose weakest part precedes, and whose strongest follows. . . Red is an impression on the Retina of an oblique and confus'd pulse of light, whose strongest part precedes and whose weakest follows.' For him blue and red were the primary colours, all others being mixed. There is no time to follow his

extraordinarily acute experiments on the colours of thin plates, but it must be noted that his ill-tempered attacks were the cause of Newton expressing a disgust with science which nearly caused him to abandon her pursuit, and were, it is almost certain, the reason why the *Opticks* was not published until 1704, the year following Hooke's death. The quarrel between the two men, both, as is evident from their correspondence, capable of generous appreciation of the other's achievements, was exacerbated by Oldenberg, then Secretary of the Society, whose dislike of Hooke may have sprung from causes little creditable to him. Newton had been undoubtedly stimulated by his reading of Hooke's *Micrographia*, and he was always very remiss in acknowledgements to him. Let us pay a tribute to poor Hooke, sickly and without position or powerful friends. He had not Newton's power of thought, but he was probably the most ingenious contriver who ever lived and was a shrewd and daring speculator.

The experiments on the composition of white light, which reduced the whole study to a quantitative basis, by showing that the refraction is a measure and index of the colour, will, as I have said, be dealt with by Lord Rayleigh. To the reflecting telescope, the explanation of the rainbow, the work on thin films, exemplified by Newton's rings, and the experiments on diffraction, this passing reference must suffice.

Nowhere more than in his writings on light does Newton stress his dislike of speculation not firmly rooted in experiment. The first words of the *Opticks* are 'My design in this Book is not to explain the Properties of Light by Hypotheses, but to propose and prove them by Reason and Experiments,' and again, in the 31st Query appended to the third edition of that book he makes his standpoint very clear, saying 'These Principles I consider not as occult Qualities, supposed to result from the specifick Forms of Things, but as general Laws of Nature, by which the Things themselves are form'd: their Truth appearing to us by Phaenomena, though their Causes be not yet discover'd. For these are manifest Qualities, and their Causes only are occult. And the *Aristotelians* gave the Name of occult Qualities not to manifest Qualities, but to such Qualities only as they supposed to lie hid in Bodies, and to be the unknown Causes of manifest Effects. . . . To tell us that every Species of Things is endow'd with an occult specifick Quality by which it acts and produces manifest Effects, is to tell us nothing.' It is in the light of this that we must read the famous 'Hypotheses non fingo'. Newton, of course, did make hypotheses, and even call them such—for instance in *Principia*, Book II, Section IX, his assumption as to the behaviour of viscous fluids is headed 'Hypothesis' and there are other instances. All he meant was that he was reluctant to speculate beyond any possibility of quantitative deduction, to form conjectures whose defence would be merely a matter of dialectics.

Newton's corpuscular hypothesis to account for the experimental behaviour of light is clearly a hypothesis, but we shall see how closely he adapts it to the observations. First, he points out that light cannot be a wave motion, or it would spread out on passing through an opening. Figure 5 is the diagram with which

he illustrates this point. He knew, it is true, the phenomena of diffraction, but he did not realize how by making the wave-length small enough this could be reconciled with the general facts of rectilinear propagation, for which a stream of particles seems best suited. He clearly saw that the phenomena of the colours of thin plates demanded a periodicity, and he introduced this into his theory by the hypothesis of fits of easy reflexion and easy transmission. It is clear that

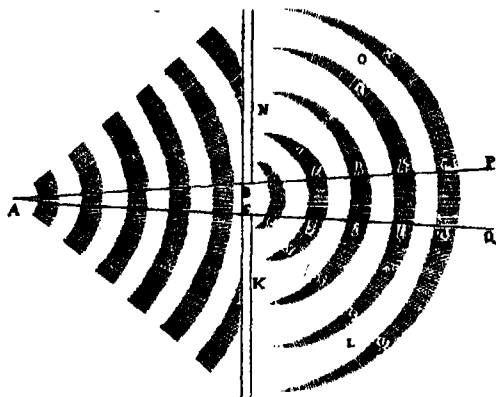


FIGURE 5

light is partly reflected and partly transmitted at the surface of a transparent body: Newton supposed that a light particle alternated at regular intervals between a state in which it was transmitted through the surface and one in which it was sent back. He put forward diffidently ('Those that are averse to assenting to any new Discoveries but such as they can explain by an Hypothesis, may for the present suppose') the idea that the impact of the particle on the surface excited vibrations in the medium which, overtaking the particles, put them into these alternating states. We are forcibly reminded of modern theories. The length of the interval of the fit, corresponding to our wave-length, was greater for the red than for the blue, and Newton gives the interval for yellow light incident normally as just about the actual wave-length of yellow light. However, of his whole theory he says 'But whether this Hypothesis be true or false I do not here consider. I content myself with the bare Discovery that the Rays of Light are by some cause or other alternately disposed to be reflected or refracted for many vicissitudes.' When he comes to consider polarization, he has to endow his particles with 'sides' so that there is a lack of complete symmetry about the direction of propagation. In short, he gave his particles just those properties which interpreted the experiment and hence was led to endow them with a periodicity and a polarity. This brought him to assume subsidiary waves accompanying the particles when they interacted with matter.

There seems to be a general belief that Huygens, as against Newton, advocated something very close to the wave theory of light, as it was accepted in, let us say, 1900. Huygens' wavelets, however, lack the essential properties with which

Fresnel afterwards endowed them: they were not only longitudinal but had what were practically particle properties, as exemplified by the fact that the pole—the point of contact with the envelope—alone was efficacious. He was just as incapable as Newton of giving a satisfactory explanation of diffraction.

If we are asked to state in a sentence what was the main effect of Newton's work on the thought of his time, I think that the answer must be that it was to establish the power and universality of the methods of quantitative science. To Galileo we owe the great service, one that cannot be too highly praised, of having made particle dynamics into a science, but he did not look beyond the earth for its efficacy nor suggest the application of his methods to the relative movement of the parts of a continuous medium, such as water. Huygens founded the study of rigid dynamics. Hooke suggested and speculated with extraordinary ingenuity and acuteness. Newton, however, showed that three clearly enunciated laws of motion applied to all observable movements of inanimate nature: they governed the motion of waves and projectiles, visible solids and invisible air, resisted as well as free movements. Together with the inverse square law they explained not only the gross movement of planets and the movement of the comets, which before had seemed capricious, but also details which nobody before had ever considered as being mechanically explicable, such as the precession of the equinoxes. The problems of the tides and of the irregularities of the moon's motion he did not fully solve, it is true, but he did enough to convince mathematicians that they were soluble by his methods. After Newton's work had been assimilated, the body of natural philosophers accepted it as a commonplace that all terrestrial and celestial movements were explicable in precise and numerical terms by calculations based on a few general laws: before Newton most thinkers were ready to invoke *ad hoc* principles and occult causes, based on human and divine analogies, for any but the simplest terrestrial phenomena, and the few who were in advance of their times were feeling tentatively for solutions which eluded their grasp.

Even in chemistry Newton was looking for an explanation in terms of attractions, though, strangely enough, in this science he never seems to have applied his own rule and made quantitative experiments. His work on light lies somewhat outside the mechanical scheme, but here again his insistence on the quantitative created a completely new attitude towards colour. It became a subject for measurement and calculation, rather than one for discussion in terms of generalities.

If we are to try to represent Newton's achievements by some modern analogy, to construct some imaginary figure who should be to our times what Newton was to his, we must credit this synthetic representative with, I think, the whole of relativity up to, and somewhat further than, the stage at present reached—we must suppose our modern Newton to have satisfactorily completed a unitary field theory. In light we must credit him both with having established the existence of spectral regularities and with their explanation in terms of the quantum theory. Possibly, too, we must give him the Rutherford atom model and its theoretical development, a simple astronomy in little to correspond to the solar system. Let us, then,

think of one man who, starting in 1900, say, had done the fundamental work of Einstein, Planck, Bohr and Schroedinger, and much of that of Rutherford, Alfred Fowler and Paschen, say, by 1930, and had then become, say, Governor of the Bank of England, besides writing two books of Hibbert lectures and spending much of his time on psychical research, to correspond with Newton's theological and mystical interests. Let such a man represent our modern Newton and think how we should regard him. Only so, I think, can we see Newton as he appeared to his contemporaries at the end of his life.

There are no discontinuities in nature and there are none in the history of science. No discovery or fundamental innovation is absolutely new, unconnected with past thought and the stirring spirit of its own time. Newton was not uninfluenced by certain of his immediate predecessors and of his contemporaries. The revolt from the introspective method of constructing explanations of heavenly and earthly phenomena by appeals to philosophic necessity had begun before his birth, and his time was rich in brilliant exponents of the experimental philosophy, whose names will always stand as stars adorning the story of science. To compare him with other men of his time and to recognize their contributions to the development of the physical sciences does not, however, lead us to think less of Newton's achievements but rather to wonder at them all the more. It is easier to estimate the size of a colossus if there are statues of more than life size in its neighbourhood than if it stands alone in a desert.

Newton owed much to the pioneer labours of Galileo, who had founded the science of mechanics, and, in a different way, much to Barrow, whose great mathematical acuteness and sympathetic support were always at the disposal of the young Cambridge scholar. To Hooke he owed more than he was ever prepared to acknowledge. From his great continental contemporaries, Huygens and Leibniz, he borrowed little, if anything. Possibly if mathematical advances were alone in question Leibniz would have to be considered as a possible rival, but for Newton mathematics were merely a means to a physical end: his mathematical innovations may even be left out of consideration without grave injury to Newton's fame. There is no record of physical experimentation that can compare for mastery and elegance with the *Opticks*, no work in exact science that produces the impression of supreme greatness and power of thought that the *Principia* does.

The spirit of this age is a denigrating one, which, in a reaction from the earlier custom of representing great men as free from all blemish and weakness, delights in attempting to show that nobody much exceeds the common level. If an earlier painter might have left out Cromwell's warts, a painter of to-day might represent his face as one huge wart. I have read record of Newton's weaknesses and I know of the adumbrations of his discoveries that can be found in forerunners and contemporaries: I acknowledge that his earlier biographer shut his eyes to any incident, writing or action that might seem to detract from his perfection. Nevertheless, all things considered, I think that the contemporary judgement of his greatness can still stand, and that, if the Marquis de l'Hôpital's query as to



ISAAC NEWTON

from a portrait by Godfrey Kneller in the possession of the Earl of Portsmouth.



*The manor house of Woolsthorpe in the parish
of Colsterworth Lincolnshire, where, Dr Isaac
Newton was born: being his own estate.*

WOOLSTHORPE MANOR

from a coloured sketch by William Stukeley, F.R.S., in his *Memoirs of Sir Isaac Newton*, a manuscript book in the possession of the Royal Society, presented by A. Hastings White, lately librarian of the Society.

whether Newton eat, drank or slept like ordinary men ('for I picture him to myself as a celestial genius') seems to our present-day sobriety an affectation, nevertheless we may agree that the line of Lucretius placed on the Trinity statue was well chosen and fitting—

Qui genus humanum ingenio superavit

'who excelled the human race in power of thought'.



The death mask.

Newton as an experimenter

BY LORD RAYLEIGH, F.R.S.

The duty has been assigned to me of telling you something about Newton as an experimentalist. As the result of a study of what is known of his history, it seems to me that among his various intellectual pursuits experiment was his first love and the love to which he was most constant. Strange though it be, he seems in some moods to have doubted whether his theoretical studies were worth while, and I do not recall any case where he expressed himself enthusiastically about them. On the other hand, he speaks of his optical work as 'The oddest if not the most considerable detection which has hitherto been made in the operation of nature.'

Newton loved the mechanical side of experimental work. As a boy he constructed sundials, and, what is more, fixed one of them into the side of the house effectually enough for it to be there a century later. A notebook of his boyhood shows him assiduous in collecting recipes for various kinds of drawing materials, and he notes methods of performing some (rather nasty) conjuring tricks. Later on, when he is making his reflecting telescope, it is obvious that he is a skilled amateur mechanic, at home in furnace operation. He builds his own brick furnace, prepares speculum metal, and is apparently more successful than the professional opticians of the time in grinding and polishing it to a satisfactory spherical figure. (The days of parabolizing were not yet.) It was not until a good many years later that they were able to put such instruments on the market.*

Asked in his old age where he got the tools for his work, he replied that he had made them himself, and could have achieved little progress without doing so.

There are occasional hints to be gleaned that Newton practised other mechanical arts. Thus, when he examines the colour of thin blown glass, it appears that he has the facilities for glass-blowing at hand, and was presumably able to use them.

So much for the base mechanical side. Newton, however, had what may be called the itch of experiment and instinctively examined in this way any natural phenomenon that excited his interest. This instinct is not a common one, and it would be of interest to investigate statistically whether it is more correlated with mathematical aptitude than with (say) an aptitude for literary and historical studies. (Newton had all these.) Although he experimented in other fields, such as mechanics, heat, and electricity, and even in anatomy and physiology, his optical experiments are of much greater importance, and in the short time at our disposal

* Among the first successful commercial makers of reflecting telescopes was James Short (1710-1788), whose instruments were of the Gregorian type. He is said to have made a considerable number of concave and convex mirrors, and to have 'married' them by trial of what pairs gave the best result.

we shall only be able to consider a part even of these. The fundamental researches on the composition of white light were read before this Society in 1872, and subsequently recapitulated in his *Opticks* (1704), and we cannot do better than concentrate our attention on them. Although the results are common property nowadays, yet on an occasion like this we shall do well to go back to Newton's own methods and point of view, and to repeat his experiments as nearly as we can in his own way.

Newton's experiments on the spectrum are sometimes presented as if he had started out with the idea of examining the composition of white light. It is true that his *Opticks* (like his *Principia*) introduces the various topics as Theorems or Problems proposed after the manner of Euclid, and most of the experiments which I hope to show you are ranged in support of his proposition that 'The light of the sun consists of rays differently refrangible'. It does not seem probable, however, that he set out in the first instance to prove this or any other proposition. He bought a prism at Stourbridge Fair (near Cambridge) in 1666, 'to try therewith the celebrated phenomena of colours'. It is clear from this that the prismatic colours were quite a well-recognized phenomenon at this time, and this is also shown by the circumstance that the art of cutting diamonds so as to display them was already long known. In Peacham's *Gentlemanly Exercises* (1612) reference is made to 'A three square cristal prisme wherein you shall perceive the blew to be outmost next to the red,' and Grimaldi and others had already experimented on the subject, though without arriving at clear views. We can readily imagine how Newton, handling the prism, would soon find that the colours were well seen in candle light, but not in diffused daylight. It would not be a long step from this to try the effect on a beam of direct sunlight admitted through a hole in the shutter.

There is no reason to think that he did this with a very clear anticipation of what the effect would be. He was exploring a nearly virgin territory. His beam of sunlight passing through the hole produced an image of the sun. It was what we now call a pinhole image, though the hole need not be very small. Newton's hole was $\frac{1}{8}$ in. diameter. Then he interposed the prism (figure 1).

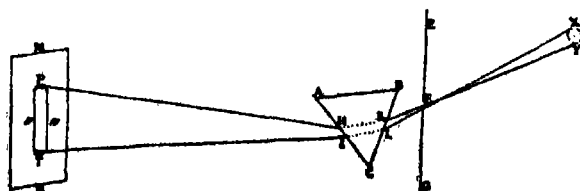


FIGURE 1

In spite of our increased control over nature, we cannot produce a beam of sunlight at pleasure late on a November afternoon, so I must be content to use the electric arc with a positive carbon presented end on, and the incandescent positive crater will represent the sun's disk. I have this in a lantern and a metal

diaphragm is placed some 2 ft. in front. (No lenses are used.) You see the image of the crater on the screen as a round disk. We will now do what Newton did, and place a glass prism in front of the hole. You see that the beam is deflected through a considerable angle towards the thick end of the prism, and you see that the image now takes the form of a band of prismatic colours on the screen. We could make this experiment more spectacular with modern resources, but my object is to follow in Newton's footsteps as nearly as circumstances allow. To most people the colours would seem the main feature of this experiment, but Newton was more surprised and impressed with the fact that the sun's image was no longer round, but was spread out into a band of which the length was some five times the breadth. The breadth of the image as projected through the prism still answered to the sun's diameter, that is to say, it was equal to the diameter of the original round image of the sun at the same distance with the prism away.

'If', he says, 'the refraction were done regularly according to one certain proportion of the sines of incidence and refraction, as is vulgarly supposed, the refracted image ought to have appeared round.'

At the same time the colours had to be taken into account. Were the colours definitely related to the differences of refraction, or was the apparent relation only incidental? In Newton's words:

'Wherever this inequality arises, whether it be that some of the incident rays are refracted more, and others less, constantly or by chance, or that one and the same ray is by refraction disturbed, shattered, dilated and as it were split and spread into many diverging rays, as Grimaldi supposes, does not yet appear by these experiments, but will appear by those that follow.'

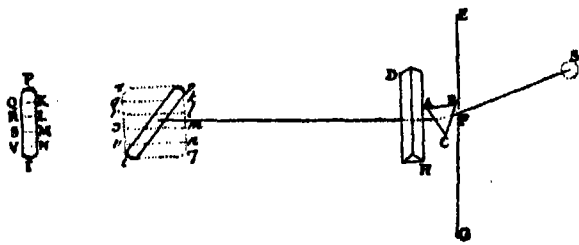


FIGURE 2

The next experiment was accordingly designed to throw light on this point. It is known as the experiment of crossed spectra (figure 2), the idea being to refract each of the coloured rays a second time, in a direction at right angles to the former, and to see whether the blue rays, which were more refracted than the red, would be so again. Also to see whether they would retain their original colour, or whether they would undergo further analysis.

We will now repeat this experiment. I put in a second prism at right angles to the former, so that the light is thrown upwards. You see that the colours remain the same and in the same order as before. The blue remains blue and the red

remains red, but the blue is again more refracted than the red, and takes a higher position on the screen, the whole image or spectrum now taking an inclined position.

Newton made several more elaborate variants of this experiment. I will show you one of them. Two holes are used, with a prism in front of each, and two spectra are then projected on the screen in line, the red of one following on the blue of the other. Refracting them upwards with a third prism, we see that this line is dislocated, and that blue and red which were before adjacent are now widely separated (figure 3).

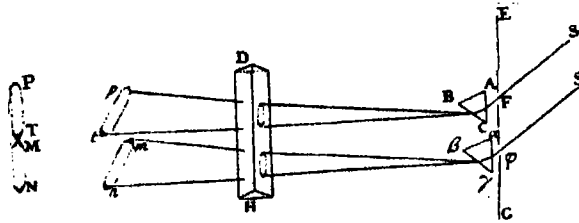


FIGURE 3

Newton in this way arrived at the view that the original white light of the sun consisted of coloured components, red, yellow, green, blue and violet. Each of these had a characteristic index of refraction and was bent to a perceptibly different extent by the prism. The result was that the original round white image of the sun was replaced by a series of coloured images side by side, overlapping and blending one with another. On introducing a second prism at right angles each of these coloured images was again refracted to its characteristic extent and each retained its colour without further analysis (figure 2).

Newton, however, soon saw the crudeness of his first arrangements, and we will now consider the methods he used to improve them. As he pointed out, the individual coloured images of the sun overlap at the middle, but not at the edges. Images of the sun are not in fact very suitable, because the angular width of the sun is about half a degree, and this is a considerable fraction of the angle between the red and violet rays. We want to make the sun in effect smaller, by blocking out part of its disk, which might be done in imagination by a distant diaphragm supported in mid-air. At the same time the hole in the shutter might with advantage be replaced by a lens, which would render the images sharper. When we have got so far in imagination, we notice that the long distance between the diaphragm in mid-air and the lens is not essential; with a suitable lens a moderate distance will do. We bring the diaphragm nearer, and fix it in the shutter, and bring the lens forward into the room, placing it so as to form a distinct picture of the diaphragm on the wall (figure 4).

We have now the diaphragm fixed in the shutter, and backed by the sun, as our effective source of light. In this respect it differs from the original diaphragm, which acted not only as a source, but as an image-forming pinhole. We now

have a picture of the diaphragm and not of the sun, and it will be advantageous to substitute a slit, which would not previously have been of any use, because the size of the image was then defined by the angular diameter of the sun. With the slit *and the lens* we shall get narrow well-defined images in each colour, and the colours will be pure and not overlapping. (Needless to say, modern refinements carry the matter much further.)

The explanation I have just given is in substance Newton's own.

We now adopt his improved arrangement, and you see the image of the slit projected on the screen. When I put in the prism, this narrow image is dispersed into a pure spectrum.

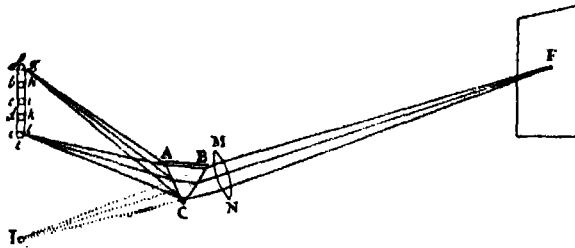


FIGURE 4



FIGURE 5

Let us now project the spectrum on to a small screen with a slit (figure 5) in it, which to avoid confusion I will call the screen-aperture. A colour passing through the screen-aperture is isolated, and you see it thrown on a second screen beyond. If I rotate the prism, I can throw the various colours in succession on the slit in the screen-aperture, so that the image on the second screen is taken through the succession of spectrum colours from red to violet. Starting with it at the red, I place a second prism in position behind the screen-aperture, and the rays are deviated, as we should expect. There is no elongation of the image this time, nor is there any diversity of colour. The red image remains red. The same applies to any other colour of the spectrum that is passed through the screen-aperture. But you will see that as the successive colours are passed, from red to violet, the deviation on the second screen continually increases, the red being the least deviated and the violet the most.

This experiment shows essentially the same thing as the experiment of crossed spectra which we saw before, but more perfectly.

The important point is that having once analysed white light into its prismatic constituent colours, this analysis is final, and further prismatic analyses can do no more. Further, each colour has its characteristic refrangibility, which will show itself at every refraction. Newton attached the greatest importance to this experiment, which he called, in Baconian phrase, the 'experimentum crucis'.

Having effected the analysis of white light into its constituent colours, he proceeded to reverse the process, and to show how the colours of the spectrum could be recomposed into white light. This he did by projecting the spectrum upon a large lens (figure 6) when the various rays were all converted to the same place, and produced a patch of white light at the focus. We will now repeat this experiment. If I place a white card immediately near the large lens, the spectrum is of course seen upon it. As we move the card away, the luminous patch contracts, and becomes white in the middle, though it remains red on the right and blue on the left. At the focus these coloured borders disappear, and there is nothing but white. Beyond the focus, where the red and blue rays have crossed, these colours are exchanged, and at longer distances still we have the spectrum inverted with the red on the left.

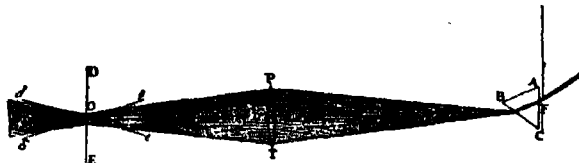


FIGURE 6

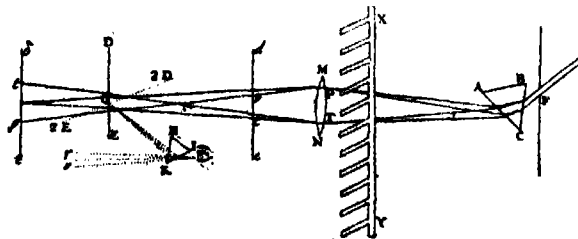


FIGURE 7

Let us now return to the focus where we have the colours converged to form white. I have here a kind of large-scale model of a comb as used by Newton (figure 7) which I can place in front of the lens. When I do so some part of the spectrum will be intercepted by the comb, and you see that the white patch produced by recombination of the entire spectrum loses its whiteness, and becomes tinged with colour. The colours produced in this way are not pure spectrum colours, because any tooth of the comb permits colour from parts of the spectrum on either side of it to pass. The colours therefore are composite, but we see that the complete spectrum is necessary to produce white. If the comb is moved slowly to and fro, you see that the coloured patch varies, being now reddish, now yellowish,

now bluish, but never white. If, however, it is moved to and fro as rapidly as may be, all the spectrum colours are present in succession. They blend by persistence of vision, and the patch becomes white again.

Newton devoted considerable attention to the colours of natural bodies, and he came to the conclusion that a coloured body is only coloured in virtue of its power of reflecting some of the components of white light more than others. I project the spectrum on a white screen. You note where the red light is seen. Now I substitute a red screen. We see red light where we saw it before, and nowhere else. The colours are not altered, though red is the only one which is strongly reflected by the red screen. The others are mostly absorbed, and their place in the spectrum is comparatively dark. Red is advantageous for such experiments, as red pigments are much the purest in colour.

In another of Newton's experiments the spectrum is projected on a large lens as before. If a white card is held at the focus, it appears white, for all the colours are converged upon it. If a card painted with cinnabar (vermilion) is placed there, it appears red as usual. If we block out the red part of the spectrum at the lens, the card shows no trace of red, becoming comparatively dark. If we block out everything but red, the card shows up with a purer red than at first.

We have now been through most of the experiments by which Newton supported his main proposition 'The light of the sun consists of rays differently refrangible,' following his own methods as closely as might be. They are a model for all time of how experimental research should be conducted, and it is difficult to our generation to see how any intelligent person could refuse his assent. This was by no means the unanimous opinion of his own contemporaries. One critic, Linus, by name, maintained that the spectrum could only be seen when the sun was shining through cloud, and it is strange, but true, that Newton was seriously discomposed by this ridiculous mare's nest. Goethe, as Professor Andrade has already mentioned, bitterly attacked Newton, using abusive expressions, and apparently thinking that to use isolated beams of sunlight admitted into a dark room was treating nature in an unfair manner, and compelling her under torture to say what she did not really mean!

The opposition to Newton's views did not finally die down till within measurable distance of our own time. In a book dated 1853, and dedicated to a man whom I as a boy personally knew, Sir David Brewster says: 'The conclusion deduced by Newton is no longer admissible as a general truth "That to the same degree of refrangibility ever belongs the same colour and to the same colour ever belongs the same degree of refrangibility"'.^{*}

Let us not make the mistake of thinking that Newton in his optical researches, exploring virgin ground, had easy triumphs. He did not escape the usual fate of discoverers. His jealous contemporary Hooke said his conclusions were not new. His admiring biographer, Brewster, concluded that they were not true!

^{*} It is not intended to criticize Brewster adversely. The appearances by which he was misled are extremely deceptive.

Newton and the science of to-day

BY SIR JAMES JEANS, F.R.S.

We are met to commemorate the greatest of our men of science, and it falls to my lot to discuss where he stands in relation to the science of to-day.

Three hundred years have elapsed since he was born, and 276 since his fertile mind had conceived most of those fundamental ideas, the development of which was to add such lustre to his name and be of such outstanding value to science—in astronomy the idea of universal gravitation; in physics the theory of colours; in mathematics the differential and integral calculus (or, as he would have said, the direct and inverse fluxions) as well as the binomial theorem and the method of infinite series. He tells us he had thought of all these before he was twenty-four years old, although he wrote little about most of them until many years later.

The intervening years have seen Newton's ideas examined, tested and developed as no other set of scientific ideas ever has been, and it might seem that by now we ought to be able to assess his greatness and place him in his rightful position relative not only to the science of to-day, but also to that of all time.

We have no doubts as to his greatness, but we probably feel less confidence in our powers to assess his ultimate position in science than we should have done fifty years ago, and certainly less than his immediate successors felt. For they were not content merely to claim outstanding greatness for Newton; they went on to claim a quality of finality and uniqueness which we know better than claim for him to-day.

Laplace, for instance, wrote that the *Principia* was assured for all time of a pre-eminence above all other productions of the human intellect, while Lagrange, after expressing himself in similar terms, went on to say that Newton was 'not only the greatest genius that had ever existed, but was also the most fortunate, for, as there is only one universe, it can fall to only one man in the world's history to interpret its laws'.

Such eulogies as these are in striking contrast with Newton's own modest estimate of himself. As his life was drawing to its close, he made the much quoted remark that to himself he seemed only like a boy playing on the sea-shore, 'and diverting myself', he said, 'in now and then finding a smoother pebble or a prettier shell than ordinary, while the great ocean of truth lay all undiscovered before me'. Which of these two estimates, we may ask, is the nearer to the truth? And how far is it possible to reconcile them? Newton's contemporaries obviously could not have reconciled them; can we?

All work in mathematics, no matter how modest it may be, may claim immortality if only it is accurate, since absolute truth can have nothing to fear from the disintegrating influence of time. But it is different with work in astronomy

or physics. This must at best be relative to the knowledge, or even to the hypotheses, of its time, so that an increase of knowledge or an abandonment of hypotheses may at any time show it to be of no permanent value. This is as true of the work of Newton as of that of lesser men.

Yet Newton's contemporaries, as well as many generations after them, made the mistake of supposing that Newton's work in astronomy and physics possessed the same sort of absolute validity as his discoveries in mathematics; they thought that no future scientist would ever be able to penetrate to the innermost recesses of Nature's temple and discover more of her fundamental mysteries, because Newton had been there already and had seen all that there was to see. But we of to-day know that Nature's temple has many chambers, room after room, and that Newton had hardly penetrated further than the ante-chamber. Our thoughts turn to the work of Planck, Rutherford and Einstein. The ante-chamber which Newton had explored was familiar ground to them, but they found the keys to other, and hitherto unsuspected, rooms; in these they discovered laws every bit as fundamental as any that Newton discovered. We see that Newton was wiser than his contemporaries in his vision of the great ocean of truth which lay all undiscovered before him. But does this add to his intellectual stature or diminish it? Should it make him seem greater or less in our eyes?

There are some—although mostly laymen in science—who see science primarily as something that is for ever changing. For them the science of any period is like the sand-castles that the children build on the sea-shore; the rising tide will soon wash them away, and leave the sands clear for the new array of castles which will be built the next day. Those who hold such views are led, somewhat naturally, to make such statements as that Newton is out-of-date and superseded.

But the comparison is obviously faulty. Science is knowledge, and the primary characteristic of knowledge is not that it is for ever changing, but that it is for ever growing. Of course, all growth implies change, so that science must continually change—like the tree which is for ever budding out in new directions—but this is only a secondary effect. The correct comparison is not with sand-castles, which change because they are continually washed away and replaced, but with a vast building which changes as one floor is built on top of another, or a new wing is built where none stood before. This building is not like a medieval cathedral, the creation of a great number of artificers, each building after his own taste and fancy. It is an embodiment of scientific truth, and the truths of science are the same, no matter who discovers them. Whatever artificers build the structure, the blue-prints have previously been drawn by Nature herself, so that, whether Newton had lived or not, the building must at some time or other have looked pretty much as it looks to-day. And if we are to estimate how much modern science owes to Newton, we must not only consider how much Newton built himself, but also by how much he expedited the building—to what extent, to put it crudely, did Newton act as a catalyst to the growth and consolidation of scientific knowledge?

Let us first consider the significance of Newton's work in pure mathematics.

There have been instances in this subject—especially in the Theory of Numbers—where a worker has stated a theorem which he believed to be true, but which centuries of discussion have not succeeded in either proving or disproving. Newton's mathematical work was not of that type. His métier was not the discovery of fanciful but comparatively useless theorems in recondite branches of pure mathematics, but the forging of practical tools which were urgently needed for further scientific progress, and we may be sure that if Newton had not forged them someone else would have done so in time.

For instance, Newton hit upon the binomial theorem in two distinct ways. He, of course, knew the expansions of $(1-x)$ raised to integral powers—squares, cubes, etc.; these are easily found by simple multiplication. Starting from these known expansions, Newton interpolated to find the values of the coefficients when $1-x$ was raised to intermediate values, such as $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, ... He also calculated the value of $(1-x)^{\frac{1}{2}}$ by taking the square root of $(1-x)$ by the simple rules of schoolboy arithmetic, and of course found that the result confirmed what he had already obtained. It is hard to imagine that such simple artifices could have remained untried for long; the binomial theorem was simply a nugget lying on the direct road of progress, and it is quite inconceivable that generations of mathematicians could have passed it by without noticing it.

It was much the same with the greater discovery of the method of fluxions. Rouse Ball tells us that, before Newton appeared on the scene at all, the method had been foreshadowed in the writings of no fewer than seven mathematicians—Napier, Kepler, Cavalieri, Pascal, Fermat, Wallis and Barrow. And in the more convenient and practical form of the differential calculus, it was actually discovered by Leibniz in 1675, and published to the world in 1684—nine years before Newton published his fluxional calculus. Apart from the personal questions involved, the bitter controversy as to whether the two discoveries were entirely independent, or whether Leibniz got ideas from seeing Newton's earlier letters on the subject, seems to me relatively unimportant; the calculus was simply another nugget lying on the direct line of scientific advance, and someone was bound to notice it and pick it up—if not to-day, then to-morrow; if not Newton, then Leibniz.

When we turn to consider Newton's still more famous work in mechanics and astronomy, we find the same story repeated; he did not drag science after him into peculiar paths of his own choosing, but kept to the broad main road, discussing the same problems as others were already discussing. The fundamental ideas with which he worked were not in any way novel, or peculiar to himself. His first law of motion had been given by Descartes in 1644, and indeed Plutarch had given it 1500 years earlier together with a foreshadowing of the general principles of planetary motions, and the conception of gravity extending as far as the orbit of the moon. In his *De facie in orbe lunae*, Plutarch had written, 'You fear that the moon may fall, on the grounds that only light air circulates under the moon and that this is not adequate to bear a solid weight. But the moon is

secured against falling by her motion and the swing of her revolution—just as objects put in slings are prevented from falling by the circular whirl. For everything is carried along by the motion natural to it, if it is not deflected by anything else.'

Much the same, again, may be said about Newton's famous law of gravitation. The attraction of lodestone for iron had familiarized men with the idea that matter could attract other matter across empty space and, as far back as 1600, Gilbert in his *de Magnete* had conjectured that the earth might attract the moon like a magnet. Kepler had gone further in the same direction in 1619; we find him insisting on the mutual attraction of all matter, instancing the tendency of all objects to fall towards the centre of the earth, as well as the ocean tides which, he maintained, were caused by the attraction of the moon. He said that two stones out in free space would approach one another, like mutually attracting magnets, and would finally meet at the point which we now call the centre of gravity of the two. He thought that the planets were kept moving in their orbits by some power inherent in the sun, and, although forming a very wrong idea of the nature of this power, he conjectured that it must vary as the inverse square of the distance. However, he immediately abandoned this law in favour of one varying inversely as the simple distance, a change which was challenged by Bouillard, who insisted that the true law must be that of the inverse square. Indeed, one of Kepler's own laws—that connecting the periodic times of the planets with their distance from the sun—shows clearly that the true law cannot be other than the inverse square law. Newton tells us that he had already noticed this in the plague years 1665 and 1666, but he did not publish it until it had been noticed independently by Hooke, Halley, Huygens and Wren.

Looking back on all this, we can hardly accept Wordsworth's description of Newton's mind as

'for ever voyaging through strange seas of thought alone'.

Except in optics, where he voyaged alone because everyone else had lost the way, Newton's ideas were very much those of his contemporaries. His main achievement was not, as is so often stated, that he was the first to think of gravity as extending to the orbit of the moon, for this had been thought of 1500 years before he was born. Neither was it that he was the first to think of universal gravitation, for many had thought of it before him. Neither was it that he was the first to conjecture that the force of gravitation must fall off as the inverse square of the distance, for this idea was generally current in his time. It was not even that he was the first to understand the general principle of planetary motions, for others understood these equally well.

His great service to science, and through science to the human race, was not that he for ever voyaged through strange seas of thought alone, but that he voyaged through familiar seas of thought—if not in company with others, at least in seas that were much frequented by others. Bacon had written of his own

contributions to scientific thought, that they did not originate in any mental powers of his own, but rather in the spirit of his age—'partus temporis potius quam ingenii'. It was this same spirit of the age that moulded the form of most of Newton's contributions to science. Scattered scientific ideas—and unproved conjectures in particular—were floating around in abundance, and lesser lights of the firmament had made small, hesitating and uncertain advances in many directions. The age was crying for a man who could systematize, synthesize and extend the whole, and it found him in superlative excellence in Newton. But beyond all this, mental powers such as Bacon had disclaimed for himself were present in overflowing measure in Newton—'qui genus humanum ingenio superavit'.

His special talents, as I see them, were: first that, out of a large mass of confused ideas provided largely by others, he was able, with his clear and acute mind, unerringly to pick out the true and discard the false; second that, having done this, his amazing mathematical ability enabled him to replace conjecture by proof, and so provide a firm basis from which he could unhesitatingly advance; third that he did advance, with quite incredible speed, sureness and directness, his supreme mathematical ability enabling him to outstrip all competitors with the utmost ease. From *conjecturing* that a thing is so to *knowing* that it is so, is usually a very long journey; it is one which Newton was so often able to take when his competitors were not. For instance, Kepler had *conjectured* that the law of gravitational force might be the law of the inverse square; Newton *knew* that it could not be anything else. Kepler was so far from certain knowledge that he immediately abandoned this conjecture for another; Newton could not be misled into so doing, because his mathematical investigations provided him with convincing proof of the truth of his ideas. And having once settled his fundamental ideas and feeling full confidence in them, he was able to stride forward into the unknown with sure and rapid steps until he had unravelled almost the whole mystery of the then known universe.

In general it was not in originating ideas but in developing them that his greatness showed itself most outstandingly; he was primarily a synthesist and systematizer. Yet here, as ever when we try to compress his genius into a formula or to map out its boundaries, an exception turns up to suggest that it overflowed all limitations; here it is the optics.

If we wish to understand the fundamental cause of his success, we cannot do better than read what Macaulay wrote in his *History*: 'In Isaac Newton two kinds of intellectual power which have little in common, and which are not often found together in a high degree of vigour, but which are nevertheless equally necessary in the most sublime departments of physics, were united as they have never been united before or since. In no other mind have the demonstrative faculty and the inductive faculty coexisted in such supreme excellence and perfect harmony.' Einstein delivers a similar judgement even more forcibly: 'In one person he combined the experimenter, the theorist, the mechanic and, not least, the artist in exposition. He stands before us, strong, certain and alone: his joy in creation and

his minute precision are evident in every word and in every figure.' And in truth he was unmatched both in the breadth and depth of his powers. Both contributed to his success, but if anything, the breadth was more in requisition, and so made the greater contribution to the success.

Yet if we are to estimate the extent to which Newton, with his outstanding mental power and unique combination of talents, expedited the growth of scientific knowledge, we cannot but place to the other side of the account his unfortunate habit of keeping his results to himself until they had been discovered independently by someone else. To some extent, this also may be ascribed to the spirit of the age. The seventeenth century had but little of our vision of science as a great benefactor to the human race, or of new knowledge as something which ought to be disclosed at once so as to expedite further progress by others. A man's scientific discoveries were his private property, with which he could do as he liked. Often his ambition was not to enable others to benefit from them, but to prevent others benefiting from them—this is why he would often publish his results in cypher. Or he might regard science, and mathematics in particular, as a sort of mental gymnasium in which contests of skill took place, and competitors issued challenges to one another to determine which of the two was the stronger or quicker. To use Newton's own metaphor, science was, more than to-day, a matter of playing with pebbles on the seashore, and seeing who could find the smoothest pebble or the prettiest shell. Thus there was less feeling of moral responsibility than now, perhaps even less interest in science for its own sake.

And so we see Newton's terrifically powerful mind playing with the problems of science like a cat which plays with a mouse and loses all interest when he has killed it, or as we play with a crossword puzzle and regard the incident as finished when we have solved it.

We can see this exemplified in nearly all Newton's major discoveries. He tells us that he had thought of the method of fluxions in the plague year of 1665. Yet before publishing it even partially (except in cypher) to the world, he allowed 28 years to elapse—years in which Leibniz published and discovered the same thing—and it was not fully published until 1736. Again in this same year of 1665, Newton satisfied himself that a force of gravity obeying an inverse square law explained the motion of the moon 'pretty nearly', and was content to leave it at that. Later, he found that this same law would result in the planets describing the exact orbits they were known actually to describe—ellipses with the sun in one focus. Again he was content to leave it there. Most men would have published this as proof positive of the law of the inverse square, but Newton kept it to himself until Halley travelled to Cambridge to consult him on this very problem. Newton then explained that he had solved the problem five years previously, but had mislaid his proof. Finally Professor Andrade has reminded us of the reluctance with which Newton published his *Principia*—probably the greatest single work of the human intellect—and of the pressure which had to be applied before he would do so.

Apart from the general secretiveness of his age, we may perhaps find three further causes for this delaying tendency in Newton: an overactive mind which refused to halt by the wayside to perfect and write out results, a disinclination to be satisfied with anything short of perfection, and an extreme sensitiveness to criticism and consequent desire to avoid controversy, this resulting more than once in a disinclination to pursue the study of science any further. In spite of these handicaps Newton bestrode the scientific world like a Colossus: what would he have been without them?

Such was the man as I see him; let us now consider how much of the scientific edifice which Newton built, or was built under his influence, is still in useful service to-day.

Outside pure mathematics, Newton's procedure was simply that of every other true man of science. He started from known facts which, in his judgement, seemed likely to be related—as for instance, if legend can be trusted, that apples fall to the ground with an acceleration of 981 c.g.s. units, and that the moon circles round the earth once every $27\frac{1}{3}$ days—and tried to represent the isolated facts of such a group as special instances of a general system of laws. If this could be done, the set of laws so discovered could be made to predict other phenomena, which could then be submitted to the test of experiment or observation. If the laws passed this test, they were assumed to correspond to something in ultimate reality; provisionally at least they were true laws, and showed how nature worked. If Newton differed from others, it was in the greater range of facts on which he worked.

In any scientific era, the search for such sets of laws is clearly restricted by the range of facts which are known, or are capable of being known, at the time. And we cannot expect scientists to discover laws of a higher order of accuracy than the observations through which the laws are discovered.

Some may challenge this, arguing that a genius of sufficient stature ought to be able to discover the true reality behind nature from quite meagre evidence—'ex pede Herculem'—and that, when this reality was known, the laws governing its behaviour could be deduced and would necessarily be true throughout the whole range of phenomena, whether known or still to be discovered. Unhappily nature is not like that. Science can, from the nature of things, never attain to a knowledge of the realities of the world, but only to a knowledge of the phenomena—of the impressions that the world makes on our senses, generally of course through instruments of precision. Our sense-organs form a sort of screen on which nature's lantern is for ever projecting pictures; we can study these pictures, but can never pass behind the screen to see how the lantern works. And as the sequence of pictures thrown on the screen is all that we can ever know or study, a set of laws that links the pictures together in a perfect systematic order is the most that science can provide, or can properly be asked to provide. If meagreness of the instrumental equipment in any age prevents the science of that age from knowing the finer details of the pictures, then we cannot expect the science of the age to provide laws governing these fine details.

For this reason we have to introduce a sort of relativity into our conception of the aims of physics; we may say that the science of any age can aim no higher than at predicting the result of any experiment that can be performed *in the age in question*. If this is conceded, we may claim that by far the greater part of Newton's work was accurate and final, relative to the age in which he lived. And this part is still of value and still in use to-day.

The enormous increase in instrumental power which has occurred since Newton's time has thrown open new fields for exploration in two directions—towards the infinitely great and towards the infinitely small, from the electron at one end of the scale to the nebulae and the whole of space at the other. Man and the man-sized world with which Newton was mainly concerned lie about midway between these two extremes.

The laws of nature are of course universal, so that the same set of laws must prevail throughout the whole of this range, but different aspects of these laws assume importance in turn in different parts of the range. So much is this the case that we may almost regard the different parts of the range as constituting separate and detached worlds in which completely different sets of laws prevail. There is the small-scale world of electrons and of atomic physics in general in which the laws of quantum mechanics prevail, the man-sized world in which the laws of molar mechanics prevail, and the world of the great nebulae in which the laws of relativity prevail. These three worlds are all governed by the same laws, but factors which are all-important in one become mere insignificant corrections in the others. Newton's work was applicable almost exclusively to the middle-sized world of molar mechanics, and so not in general to the other two worlds, the existence of which was hardly suspected in his day. Lagrange was wrong in thinking that there was only one universe, and that it could fall to only one man to interpret its laws; there are worlds within worlds, and Newton had only interpreted the laws of one of these. Relative to this world most of Newton's work was accurate and final, which means that it was also accurate and final relative to the age in which he lived.

If we examine any recent paper on theoretical physics, we are likely to find it plentifully sprinkled with either the symbol h or the symbol c , or both. These two symbols were unknown to Newton, at least in the sense in which we now use them; they are, so to say, the emblems of the theories of quanta and relativity which did not invade physics until the present century.

The symbol c denotes the velocity of light, which is greater than the velocity of any material object, and is usually enormously greater. Also, in dynamical problems the ratio of these two velocities can only enter through its square. We are now familiar with electrons which move almost as fast as light, but the kind of dynamics with which Newton was concerned contemplated no velocity greater than that of the planet Mercury, which is about 48 km. a second. For this, the ratio of the squares of the two velocities is 0.0000000256, or, to seven significant figures, is *nil*. If we disregard this fraction, the dynamics of relativity becomes

absolutely identical with that of Newton, so that to seven significant figures Newton's dynamics was adequate and factually perfect for an age in which nothing moved faster than the planet Mercury.

It is the same with the h of the theory of quanta. This measures the atom of action which Planck discovered in 1899, and although it is of the utmost importance to phenomena on the atomic and subatomic scale, it is absurdly small in comparison with the amounts of action involved in the activities of everyday life. Roughly speaking, the ratio is that of an atom to a gramme, or of a gramme to a star. This small quantity again was quite negligible in all the problems with which Newton was concerned.

Thus the Newtonian mechanics, dynamics and astronomy were absolutely right factually, except for quantities which were so small that they could not either affect Newton's problems in any way, or make their existence known through the best instrumental accuracy that was available in his day. To see Newton's work in these subjects in their proper perspective, we must think of him as the builder of the ground floor of the great building of mechanics, dynamics and astronomy, and as its architect for many floors above this. Newton's first and last serious contribution was his *Principia*, but after this floor after floor was added on strictly Newtonian lines—on the dynamical side, the generalized dynamics of Lagrange, Hamilton and Jacobi; on the astronomical side, the dynamical astronomy of Laplace and many others; on the physical side, the electrodynamical theory which we associate primarily with Clerk Maxwell. All this and much more followed the Newtonian architectural plan without any deviation whatever for two centuries. It was not until 1887—just 200 years after the appearance of the *Principia*—that the Michelson-Morley experiment first showed that the Newtonian scheme was in actual fact imperfect. But the experiment had to be discussed for eighteen years before this inference was drawn, and it was only when Einstein's restricted theory of relativity appeared that the Newtonian scheme was generally recognized to be inadequate.

It is, however, inadequate only with reference to the ultra-refinements of modern science. When the astronomer wishes to prepare his *Nautical Almanac*, or to discuss the motions of the planets, he uses the Newtonian scheme almost exclusively. And the engineer who is building a bridge or a ship or a locomotive does precisely what he would have done had Newton's scheme never been proved inadequate. The same is true of the electrical engineer, whether he is mending a telephone or designing a power-station. The science of everyday life is still wholly Newtonian. And it is impossible to estimate how much this science owes to Newton's clear and penetrating mind having set it on the right road, and this so firmly and convincingly that none who understood his methods could doubt their rightness. Without his guidance, smaller men might have argued for centuries as to what was right and what was wrong. When Newton had spoken, it soon came to be recognized that the time for controversy as to which was the right road was past; it only remained to advance along the Newtonian road.

Newton's work in optics stands somewhat apart from the rest in that he was much less indebted to his predecessors and contemporaries. Indeed, current ideas on the nature of light were so vague and erroneous that he was probably hindered rather than helped by what he knew of them, and was perhaps fortunate in not knowing more. We know that he attended the lectures of Barrow in Cambridge, where he would doubtless become acquainted with the optical theories of Descartes as well as with those of Barrow himself. Now Descartes thought that all space was filled with substance, and that vision resulted from the transmission of a pressure from one particle to another of this substance—much as a blind man pokes about with a stick. He thought that different sensations of colour resulted from the particles rotating with different speeds, red resulting from the most rapid rotations, and yellow, green and blue following in this order. Barrow held even more fantastic opinions; he thought that red light was more concentrated than light of other colours, but was broken up by intervals of complete darkness, yellow light consisted of a mixture of this red light with white, and so on. Hooke conjectured that light consists of a rapid vibrational motion of an ethereal medium filling all space, different colours being produced by different shapes of wave; blue, for instance, resulted from 'an oblique and confused pulse of light'.

Into the midst of all this confusion of thought came Newton, to whom, in Einstein's words, 'Nature was an open book, whose letters he could read without effort. The conceptions which he used to reduce the material of experience to order seemed to flow spontaneously from experience itself, from the beautiful experiments which he arranged in order like playthings.' With the prism he had bought at Stourbridge Fair, he went at once to the crucial experiment of the spectrum. Finding that this was longer than it was broad, he at once saw that differences of colour must result from different degrees of refrangibility. This discovery alone, the subject of Newton's first scientific paper, published in our *Philosophical Transactions* of 1672, at once took the theory of light leagues beyond the point to which any previous investigator had brought it, and opened the road to further rapid progress.

But Newton brought his views into disrepute in the eyes of many by appearing to think of light as a material substance—Hooke definitely brought this charge against him. Actually Newton repeatedly claimed that his results did not depend on any special views as to the ultimate nature of light, and he was careful to avoid such words as 'corpuscles' which might seem to imply such views. Yet much of his thought seems to become meaningless unless we identify his rays—the 'least parts' of light—with something of a corpuscular nature. However little he may have said about it, Newton seems always to have had his own picture of the structure of light before his eyes, and this picture was corpuscular.

For two centuries many thought of this as the one serious mistake that Newton ever made. Recently the pendulum has swung—perhaps too far—in the other direction. Light is a transfer of energy from matter at one place to matter at another, and this transfer can only take place by complete units, or 'quanta'.

This led Einstein to picture light as a shower of arrows, each arrow containing a single complete quantum of energy, and some have hailed Newton's corpuscular theory of light as a brilliant prevision of this.

But we must remember that Newton introduced his corpuscles primarily to explain the rectilinear propagation of light, whereas Einstein introduced his light-arrows to picture the quite different property of atomicity. This being so, the similarity of the two pictures can hardly be quoted as a supreme example of intuition on the part of Newton. It might be if light really did consist of corpuscles; Newton would then have arrived at the truth by one road 200 years before Einstein reached it by another. But no competent physicist can for a moment suppose that light consists either of corpuscles or of light-arrows; these are only pictures we draw for ourselves to help us understand parts of a truth which we know, as a whole, to be for ever beyond our comprehension.

At this point we are inevitably brought to consider Newton as a philosopher. Unlike his great predecessor Descartes and his great contemporary Leibniz, he seems to have had very little interest in philosophy for its own sake. He was what we might describe to-day as a pragmatist; his main concern with a philosophy was not as to whether it was true, but whether it worked and yielded useful results. It was another illustration of the ever-recurring difference between the British and the Continentals.

Descartes had taught mathematicians to measure co-ordinates from fixed lines on a piece of paper. Newton wanted to be able to measure mechanical and astronomical co-ordinates in the same way from positions in space and instants in time. So he assumed without further ado that time could be measured from a fixed standard instant, and that the remotest parts of the universe contained vast immovable masses, providing standard positions against which absolute motion could be measured.

This assumption of the existence of absolute time and space was a mere working hypothesis which, as he admitted, might not be true at all: 'it may be that there is no body really at rest, to which the places and motions of other bodies can be referred'. But the assumption, whether true or false, produced results—nothing less than a consistent scheme which explained a large part of the then known workings of nature. This being so, philosophers such as Bruno and Leibniz might argue in vain that absolute space, time and motion were meaningless and illogical.

It was much the same in mechanics. In the man-sized world of seventeenth-century science, motion was the result of pushes or pulls of the kind that a horse exerts on a cart. Other suggested causes of motion, such as action at a distance, were regarded much as we of to-day regard levitation or table-turning, and described as 'occult qualities'. It was inevitable that Newton's force of gravity should come under this description. Leibniz wrote 'Some men begin to revive, under the specious name of forces, the occult qualities of scholasticism, but they bring us back again into the Kingdom of Darkness'. Newton's reply was a fine defence of his pragmatistical philosophy. To understand the motions of the planets under

gravity, he says, even though we do not know the cause of gravity, is as good progress in philosophy as is a knowledge of the motions of the wheels of a clock in the philosophy of clockwork, even though we do not understand why the weight which moves the wheels falls earthward.

At the same time, he refused to admit that his gravitational forces were not wholly mechanical. From the celestial phenomena, he said, we calculate the force of gravity with which bodies are drawn to the sun and the several planets. Once having found these, we deduce the motions of the planets, the comets, the moon and the sun, and 'I wish we could derive the rest of the phenomena of nature by the same kind of reasoning from mechanical principles'.

It is interesting to speculate as to what might have happened if Newton had shared the views of Leibniz as to the relativity of all motion. It is conceivable that he might have arrived speedily at what we now call the restricted theory of relativity, that through it he would have seen the impossibility of action at a distance, and would then have developed a relativity theory of gravitation similar to that of Einstein. It is conceivable, but I can hardly think it would have happened—the necessary physical ideas and mathematical technique seem to be centuries removed from the fields in which Newton worked. It is more likely that he would have been compelled to come down to an approximation which would have brought him back to the Newtonian mechanics and the inverse square law. If so, his erroneous philosophical background of absolute space and time was a real advantage, since it provided him with a short cut to a region which would have had to be explored anyhow as a preliminary to further progress. It enabled him to go direct to the main point as it appeared in the seventeenth century, neither hindered by any philosophical subtleties nor haunted by any fears that his system might not be an expression of final absolute truth.

In this way Newton created a system which, although philosophically unsound, was destined to endure for two centuries before any chinks were found in its armour. Then more philosophically minded Continentals had to be called in to patch up the deficiencies.

In this quality of practical and opportunist directness, as in many others, Newton seems to me to be typically British. In general he was strong where the British are strong, and weak where the British are weak. If we think of his various pre-eminences in turn, and consider who runs him closest in each, we shall find it is usually a British scientist who comes to mind. If we try to match the simple directness and the clear logical sequence of his experiments in optics, our thoughts turn to our own Faraday and Rutherford. Clerk Maxwell runs second to him, I think, in other aspects of his work. But if we think of those who excelled him in his philosophical outlook, we think exclusively of Continentals, from Leibniz on. His successes were typically British, and his failures—in so far as he ever failed—were also typically British. We can not only be proud that this greatest of all scientists was our own countryman, but we may reflect with satisfaction that he was no freak genius or sport, but was the concentrated embodiment of all the distinguishing characteristics of British science.



ISAAC NEWTON

from a portrait by Johan Vanderbank in the possession of the Royal Society,
presented by Martin Folkes, P.R.S.



REFLECTING TELESCOPE

made by Isaac Newton with his own hands, 1671. Presented to the Royal Society by Messrs Heath and Wing, 1766.



SOLAR DIAL

made by Isaac Newton when a boy, taken from the wall of Woolsthorpe Manor House. Presented to the Royal Society by C. Turnor, F.R.S.

Exhibits

By permission of the Managers the three lectures printed above were given in the Royal Institution, where exhibits illustrating some of Newton's experiments, arranged by Lord Rayleigh, and a small collection of Newtoniana were also on view. The programmes are reprinted below.

EXHIBITS ILLUSTRATING SOME OF NEWTON'S EXPERIMENTS

ARRANGED BY LORD RAYLEIGH

as a supplement to those shown in his Lecture on

Newton as an Experimenter

EXHIBIT No. I. *Opticks*. Book I. Experiment I. To show that lights which differ in colour differ also in refrangibility.

Newton's account abbreviated:

I took a black oblong stiff paper. One part I painted into a red Colour, the other into a blew. This paper I viewed through a Prism of Solid Glass. I found that if the refracting Angle of the Prism be turned upwards so that the paper may seem to be lifted upwards by the Refraction, its blew half will be lifted higher by the Refraction than its red half. But if the refracting Angle of the Prism be turned downwards, so that the Paper may seem to be carried lower by the Refraction, its blew half will be carried something lower than its red half.

EXHIBIT No. II. *A* and *B*. *A. Opticks*. Book I. Part II. Experiment XVI. In this experiment a uniform field of light is viewed through a reflecting prism. Newton used the cloudy sky. An illuminated opal glass is here substituted. The bright part of the field is the region of total reflexion. The relatively dark part of the field, corresponding to steeper angles of incidence, is the region of partial transmission and partial reflexion. Observe the blue arc marking the limit of total reflexion. This shows that white light contains a constituent (blue) which begins to be totally reflected at angles of incidence less steep than are necessary for the other components.

This blue colour being made by nothing else than by reflexion of a specular superficies seems so odd a Phenomenon, and so unaccountable for by the vulgar Hypothesis of Philosophers that I could not but think it deserved to be taken notice of.

B. A supplement is arranged to this exhibit, allowing the transmitted light to be examined. A pair of right-angled prisms is used forming a cube after the manner of Newton, reflexion occurring at the diagonal interface. The field is divided into a bright and a dark part as before, the dark part (no transmission) being the area of total reflexion. A reddish yellow arc marks the limit of transmission, which is complementary to the blue arc seen by reflexion.

This experiment was performed by Newton in a more elaborate way, using the sun's light and projecting a spectrum of the beam transmitted by the cube. The present arrangement is adopted to avoid the complication of an arc lamp, and reproduces Newton's essential point.

EXHIBIT No. III. Opticks. Book II. Part I. Observation xiii. This experiment shows that Newton's rings, formed between a spherical surface of large radius and a flat plate, are larger in red light than in blue. By pulling a string the observer may rotate the spectrum-forming prism and observe the rings dilating as the colour is changed from blue or blue green towards red. On releasing the string the colour passes back towards blue and the rings contract again.

Appointing an assistant to move the Prism to and fro about its Axis, that all the Colours might successively fall on that part of the Paper which I saw by reflexion from that part of the Glasses where the Circles appeared, so that all the Colours might be successively reflected from the Circles to my Eye whilst I held it immoveable, I found the Circles which the red light made to be manifestly bigger than those which were made by the blue and violet. And it was very pleasant to see them gradually swell or contract according as the Colour of the Light was changed.

Newton recognized that this proved that there was a coarser structure associated with red light than with blue. In Query 13 at the end he says:

Do not several sorts of rays make vibrations of several bignesses, which according to their bignesses excite sensations of several colours?

EXHIBIT No. IV. A, B and C. This is an experiment of Newton's on frictional electrification. (Description below.) Rub the upper surface of the glass with flannel, or in the case of *C* stroke it with the dry hand; and the movements of the pith or paper fragments will be observed. *C* is the easiest and should be tried first.

A. An apparatus with flint glass as used by Newton. The action of this is uncertain, depending on the hygroscopic condition of the glass. It may fail, as it did when first tried at the Royal Society. Application had to be made to Newton for further directions. In dry weather and with vigorous rubbing it works admirably. Try reversing the glass and rubbing first one side, then the other.

B. An apparatus with silica glass. This can be taken out and warmed over a spirit lamp. It will then work without fail.

C. An apparatus with a sheet of celluloid replacing the glass. This will work at any time without fail, if lightly stroked with the dry hand.

From the Minutes of the Royal Society. 9 December 1675.

That [Newton] having laid upon a table a round piece of glass about two inches broad, in a brass ring, so that the glass might be about one-third of an inch from the table, and the air between them inclosed upon all sides after the manner as if he had whelved a little sieve upon the table; and then rubbing the glass briskly, till some little fragments of paper, laid on the table under the glass, began to be attracted and move nimbly to and fro; after he had done rubbing the glass, the papers would continue a pretty while in various motions; sometimes leaping up to the glass and resting there awhile; then leaping down and resting there, and then leaping up and down again; and this sometimes in lines perpendicular to the table sometimes in oblique ones; sometimes also leaping up in one arch and down in another divers times together, without sensible resting between; sometimes skip in a bow from one part of the glass to another, without touching the table; and sometimes hang by a corner and turn often about very nimbly, as if they had been carried about in the midst of a whirlwind; and he otherwise variously moved every paper with a diverse motion. And upon sliding his finger on the upper side of the glass, though neither the glass nor inclosed air below were moved thereby, yet would

the papers, as they hung under the glass receive some new motion inclining this or that way, according as he moved his finger.

The experiment he proposes to be varied with a larger glass placed farther from the table, and to make use of bits of leaf gold instead of papers, esteeming that this will succeed much better, so as perhaps to make the gold rise and fall in spiral lines or whirl for a time in the air, without touching the table or glass.

Ordered that this experiment be tried the next meeting.

EXHIBITS ARRANGED BY THE ROYAL SOCIETY

1. The manuscript of the *PRINCIPIA*, from which the first edition was printed, with corrections by Newton.
2. The *REFLECTING TELESCOPE* made by Newton with his own hands in 1671.
3. A *SOLAR SUNDIAL* made by Newton when a boy. It was formerly in the wall of his home, the Manor House, Woolsthorpe.
4. A testimonial given by Newton in favour of William Jones (later F.R.S.), an applicant for a post on the staff of Christ's Hospital.
5. An *ASSAY FURNACE* used by Newton in His Majesty's Mint in the Tower of London for his experiments in cupellation.
6. A plan of His Majesty's Mint in the Tower of London, drawn for Newton on his appointment as Master Worker.* It is dated January 1701.
7. A holograph draft of Newton's report on *The value of gold in proportion to silver in several parts of Europe*.
8. Nine letters from Halley to Newton relating to the publication of the first edition of the *Principia*. They were written between 22 May 1686 and 5 July 1687.
9. *The hunting of ye green lyon* and *The standing of ye glass for ye time of putrefaction & congelation of ye medicine*. Alchemical verses copied by Newton with his own hand.
10. An *IVORY BUST* by Cheverton from the sculpture by Roubiliac.
11. *THE MANOR HOUSE OF WOOLSTHORPE*. A drawing in *Memoirs of Sir Isaac Newton's Life*, written by William Stukeley in 1752.
12. The mask of Newton's face, which belonged to Roubiliac, from the cast taken after death.

* The Deputy Master of the Mint has supplied the following note: 'In the 18th century the Royal Mint operated, in form, under the constitution laid down by Edward I. The Master Worker was a salaried officer of the Crown who contracted for the execution of the coinage, in law for the whole, in practice for a large part of the work. The Warden was a check on some of his actions, the Comptroller on others. Sir Isaac Newton was Warden from March 1696 to December 1699 and Master Worker from December 1699 to December 1726.'

ACKNOWLEDGMENTS

The President, Council and Fellows of the Royal Society wish to express their gratitude to Professor E. N. da C. Andrade, Sir James Jeans and the Lord Rayleigh, who gave the lectures: to the Managers of the Royal Institution for the use of their lecture room and apartments, as well as to the staff of the Institution for co-operation and assistance.

The Royal Society also gratefully acknowledges the courtesy of those who have given permission and facilities for the reproduction of illustrations, and offer their thanks:

To Mr H. M. Hake, C.B.E., Director of the National Portrait Gallery, for a portrait of Newton by Vanderbank and advice on other portraits;

To the Earl of Portsmouth, for the portrait of Newton by Kneller;

To the proprietors of Country Life, who own the negative of the photograph of the Kneller portrait;

To Mr Walter Lee of Grantham, for the two photographs of Woolsthorpe Manor.

The Society is also indebted to the Deputy Master of His Majesty's Mint for the loan of an assay furnace used by Newton, and other exhibits; to the Lord Keynes for the loan of manuscript letters and verses; and to the Astronomer-Royal for the loan of an ivory bust of Newton.

FERRIER LECTURE

Fatigue following highly skilled work

By F. C. BARTLETT, F.R.S.

(Delivered 29 May 1941—Received 29 May 1941)

If the character and causation of fatigue following highly skilled work are to be understood, the first need is for the discovery of more relevant and experimentally controlled facts. Unfortunately, almost all the investigators who have attempted to study fatigue of this type have adopted methods taken over with very slight change from those which have proved valuable in the study of simple muscular fatigue. They have chosen elementary operations usually considered to require some 'mental' effort—such as easy calculations, word or colour recognition and naming and the like—have repeated these operations over and over again for long periods, and have tried to express the resulting fatigue in terms of the diminution in quantity or quality of the work done. The skill fatigue of daily life is not set up under such conditions. Routine repetition of simple actions is not a characteristic of any highly skilled work, and least of all of work having a strong 'mental' component. The operations involved here are marked by complex, co-ordinated and accurately *timed* activities. The stimuli in response to which these activities are set up are neither simple nor do they usually fall into an order of fixed succession. They have the character of a field, or a pattern, which has become very highly organized, and may retain its identity in spite of a great diversity of internal arrangement.

It is possible to develop fully controlled experimental situations in which these realistic considerations have full play. When this is done the picture of fatigue following highly skilled work which emerges has certain strongly marked characters.

In such fatigue the 'standards' accepted and followed by the central nervous system unwittingly deteriorate. The operator tends to think that he is doing better work, because errors treated as significant all the time get wider and wider limits. Until a stage of great fatigue is reached, it is far more likely that the right actions will be performed at the wrong times than that the wrong actions will be performed. If accurate timing is insisted upon, gross mistakes of action may appear. The stimulus fields split up. Its pattern character alters. It becomes a collection of unconnected signals for action, with some of these predominant over all the others. Particularly, stimuli which are in the margin of the pattern, not closely organized with the central field, are ignored, 'forgotten', and serious lapses of specific reactions occur. There is a marked change in the effect of certain 'distracting', or additional stimuli. Sensations of bodily origin, in particular, become more pressing and insistent and affect the performance in ways peculiar to the tired operator. Side by side with all these changes go constant subjective symptoms. Verbal reports about any circumstances connected with known failure of performance become increasingly inaccurate, and errors are regularly projected upon objective conditions, or attributed to the interference of other people. There is a tremendous growth of irritability.

An attempt is made to discuss the light thrown by this picture upon the relation of high-level central nervous functions to simpler neuro-muscular mechanisms.

When Sir David Ferrier published his book *The functions of the brain*, he wished to find a brief statement which would best express the spirit of his work. He chose a passage from the writings of G. H. Lewes: 'Indeed every discovery is a verified hypothesis; and there is no discovery until verification has been gained: up to this point it was a guess, which might have been erroneous.' It is, I hope, precisely in such a spirit that I offer this lecture in memory of a great neurologist.

There can be few subjects more full of guesses than that which has been con-

cerned with a study of those types of fatigue which involve complex activities of the central nervous system. In no direction have these guesses been more common, or more unfortunate, than in the determination of methods of study. Over and over again, as experiment has advanced to a stage in which it must tackle complex biological problems, investigators have felt themselves bound to adopt, with little change, the methods which have won deserved success in a study of parallel, but very much simpler, questions. As everybody knows, muscular fatigue has been brilliantly and fruitfully studied by experimental methods in which isolated muscle, or relatively isolated small groups of muscle, have been subjected for varying periods to a succession of exactly similar stimuli. The resulting contractions have been recorded and analysed in many ways, but the essence of the method is always the same: it is to produce long series of responses in local mechanisms in the course of which the same reactions are repeated over and over again, though, perhaps, with diminishing efficiency.

When people began to try to apply experiment to the study of tiredness produced by the exercise of skill, they seem to have lost forthwith all their capacity for realistic imagination. They thought, and have continued to think, almost solely in terms of exactly repeated activity. Thus Kraepelin—who did far more harm than good, I think, in this field—introduced his method of the reckoning test: a person must add, or subtract, or multiply, or divide simple digits for hours on end. Others have embroidered on the same theme. They have exposed colours, or words, or easy drawings in long series, asking all the time for recognition, or naming, or both. They have required letters, or letter combinations, words or word combinations, to be written, or perhaps to be cancelled, over and over again. They have set their victims to copy line diagrams, or designs, so often that one would think that all but the most faithful must have been reduced to profanity or to tears, at least through boredom if not through fatigue. Through it all runs one great, unverified guess: fatigue must consist of diminished efficiency of specific performance due to the repetition of that performance.

Suppose, instead of rather blindly taking over methods which were just and correct when applied to the case of simple muscular fatigue, that people had honestly asked themselves what looks to be the character of the skills involved when we say we get tired in the pursuit of complex activities in daily life, in industry, or in the practice of specialized skill in the fighting services. They would have got a picture wholly different from that of repeated movements set up in response to recurrent and unvarying stimuli. They would have found co-ordinated actions the constituents of which can, and frequently do, change places. They would have found a type of behaviour in which it has become of enormous importance to *time* the constituents correctly, so that each can flow readily into its neighbour. They would have found interruptions and rests no more uniform than actions. They would have seen stimuli to such co-ordinated action which are not a repetitive succession but a field, a pattern, an organized group of signals capable of changing their internal arrangement without loss of their identity as an organized group.

This is the situation which must be brought within full experimental control if the tiredness, or fatigue, or strain following highly skilled work is to be understood.

It has proved possible to do this, but for various reasons I cannot now so describe the setting designed to investigate these fatigue effects that it may be exactly identified, for it had a very strong and realistic interest, bearing on current affairs.

The subjects of the experiment sat comfortably, surrounded by instrumental controls and faced by a panel containing all the main signals for action. These signals were in three chief groups: in the middle a group important throughout the whole period of the test; to one side a group important only at certain stages, especially at the beginning and the end; to the other side a group intermittently important indicating occurrences calling for prompt, but only occasional, action. Above and below were stimuli which could be brought in at the experimenter's will, each calling for a specific response, but still less closely bound up with the central task.

To all of these signals the operator must respond by co-ordinated movements of hands, feet, eyes and understanding. Moreover, with the exception of the very occasional stimuli, the significance of no signal was detached or isolated from that of any other. All the main stimuli were of the common form, for instrumental maintenance or control, of pointers moving across dial faces at different speeds and in different directions. Some of the subordinate stimuli were lights of varied colour. Each stimulus had a determinate relation to every other, so that both how and when any happened were by no means matters of chance. Also the response required itself helped to determine the next following pattern of stimuli, just as it usually does in real life. Consequently the skilled operator did not need to examine the signals one by one, in a regular order, but could take his cues from the whole pattern, with certain constituents always playing the dominant parts, again as happens in the skill of daily life. Since all stimuli and all reactions were moving in determinate relationships, the whole instrument was after the style of a rather complex calculating machine, and it became possible to record accurately the amount, direction and, most important of all—as it turned out—the timing of the essential elements in the operator's skill.

There is one thing more. At the beginning of the experiment this was what Ferrier would certainly, and correctly, have called an unverified guess. If one looks fairly at the tiredness which follows skill it is hard to resist the view that abstinences are as important as performances. When work begins, and the central nervous system is alert, keen and high in vigilance, its inhibitory activities are in perfect trim. Additional, irrelevant, unwanted and distracting stimuli are not within the effective field at all. But as work continues and fatigue grows, the inhibitions perhaps relax, until the skill-tired man is doing, not less work, but more, much more.

The instrumental setting was so designed that when the operator used his hands and feet to manipulate levers and switches, his eyes to follow moving pointers and

flashing lights, he could—and in fact he did—interpret the result, not in terms of revolving wheels, engaging cogs, shifting currents, vibrating rods and pistons, but, just as any man does when he drives a motor car, in terms of the whole machine. The machine, though in fact stationary all the time, seemed as if it could move, change direction, vibrate, and no single skilled operator of the men who have attempted the task has failed to experience the effect of his skill as a total, combined performance of the machine.

When a man moves with a machine there are inevitably set up a mass of proprioceptive impulses coming from changes of stress and tension at various points of the body surface, from underlying tissues and from other bodily sources. These are not original conditions of the acquisition of the skill concerned, but they are inevitable additional stimuli occurring while the skill is being acquired and when it has been mastered. Sometimes they have the character of distractions, and it may be that the fresh and the tired nervous system react differently towards them. They conform to the principle which ought to be observed in all experimental study of additional or distracting stimuli: they are an integral part of the situation itself, and not merely disconnected signals put in at the whim of an experimenter. The neglect of this principle has led to much unfortunate misunderstanding about the functions and manner of operation of distractions. In the experiment as designed it was not difficult to introduce the required proprioceptive impulses at the appropriate points. The guess was that they might act differently as strain, or fatigue, increased.

Such was the experimental setting. There was little or nothing sheerly repetitive in the stimuli or the behaviour, and nothing at all demanding strong physical effort. The task was one requiring great skill and, as it proved, of absorbing interest. That the experiment was possible at all was due mainly to the brilliant mechanical inventiveness of Dr K. J. W. Craik, and that noteworthy results were achieved was due to the patience and high experimental competence of Mr G. C. Drew, both of these being members of the Psychological Department of the University of Cambridge.

I turn now to a description of the main results. The picture which emerges may, I think, fairly be said to be the first reasonably complete representation that has been drawn of the fatigue following highly skilled work. All the phenomena to be described were obtained from a large group of operators. They are statistically reliable to a high degree. There is less than one chance in a hundred that they represent mere accidental occurrences.

We may, perhaps, see the beginning of skill when nervous mechanisms acquire the character of 'graded' response, adjusting the amount of their action to the variable intensity of the stimulus. As skill develops into concerted action, where all subordinate constituents are organized around a few outstanding ones, a new kind of 'grading' appears. Effective stimuli now acquire an 'indifference range' within which stimulus changes, though they may be appreciated, do not call for compensating activity. Variations outside that range at once set central control awake and

compensating reactions appear. In the experimental setting two of the instruments were generally central, or dominant. One recorded the extent of deviation of the machine from any given course and the other the speed of movement in any direction. At the beginning the operator, watching the pointer which showed extent of deviation, allowed it to move two or three degrees on either side of the vertical and then began at once to make the necessary control reactions. A little later he was letting it move five degrees in either way, then ten degrees, and finally it could swing from side to side over a wide range before anything was done. Exactly the same happened in the case of the speed indicator, and in fact with every other important recording device in the total group of stimuli, I think that everybody who drives a motor car will be able to observe just this same type of widening indifference range of stimuli as he becomes more and more tired.

One way of putting this is to say that skill fatigue is marked by a progressive lowering of standards of performance. But if this is so the standard is certainly not normally formulated, not one of which the operator need be aware. It is a truly physiological function of the central nervous system. It means that at any moment in the continuous exercise of skill every leading stimulus has a range of variation within which the central nervous system exercises no direct and overt control over the local responding mechanisms, and that this range has no fixed limits, but increases progressively as work continues. A large majority of the experimental operators finished their task satisfied that they had improved steadily all the time, though in fact the increase in large errors over small ones, in the case of deviation from a wanted course was 400 % as between the first and last periods of a two-hour experimental run, and 92 % in the case of speed control. Actually, in relation to their unformulated standards, they *were* doing as well at the end of their task as at the beginning, and perhaps even better.

Now suppose that one element of the response, the simple control, for example, of direction or speed, were taken out of its skill setting and required merely as an element at any stage of the experiment from beginning to end. It showed no deterioration. The absolute efficiency for the element remained unaffected. It is therefore no wonder that an experimental method which has taken elements out of skill has produced results which are strikingly at variance with everyday observation. It is not the local response that has lost its accuracy or its power. It is the central control which has functionally, but without knowledge, expanded the limits of its indifference range.

We can go farther than this with our analysis of the large errors which bunch up as fatigue increases. It will be recalled that the operator was faced with a group of changing signals from which he must select the most significant ones and time his response to these correctly. An instrument may, for example, record an unwanted change of direction. Then the operator may use a control which accelerates that very change, and if he does he has performed the wrong action at the right time. Or he may make delayed use of a control which neutralizes this unwanted change. Or yet again he may manipulate a control which itself produces a consequential

change of direction, and then bring in the counteracting control too soon. That is, he can do the right actions at the wrong time, either too late or too soon.

The overall results of the experiment showed a non-significant increase in wrong actions done at the right time, but a highly significant increase in right actions done at the wrong time. And this same type of deterioration was shown in another way as well. The whole task had to be performed to a time schedule, and to help the operator to keep to this he had a clock in his instrument panel and a stopwatch at his side. As the task proceeded the clock and the watch faded out and might almost as well not have been there at all. Of the result the operators were, apparently, faintly aware, for most of them said that their timing might have become a little erratic at the end. In fact the time estimates made by them could be as much as 200 % wrong.

So much for some of the facts concerning the deterioration of the skill response. Now let us see what happened as regards the organization of the group of stimuli. It has already been said that all the instruments on the panel were related in a determinate manner. When an operator was fresh a glance at the dominant signals meant an interpretation of the whole panel, and a movement of a controlling lever meant something that the machine was doing, or would very soon begin to do. As the task continued the panel split up, so that it became twenty or so separate recording instruments. And the controlling movements split up also, so that when any one was made it was not pictured in a pattern of machine control, but only as the correction of a particular instrument reading.

When this stage was reached it appeared as if central nervous control tended to slacken either momentarily or finally. The operator's temperament surged up and took charge of his behaviour. For the reaction to one instrument only, when reaction to many was called for, generally set a lot of the others beyond their limits of indifference range. Either there was a moment of flurry and then central control reasserted itself, or there was panic and the operator, dashing from one control to another, pulled his machine to disaster.

The splitting of the stimulus field and the corresponding progressive dissociation of actions did not occur in a haphazard manner. The field contained a central sector important all the time, side sectors important at specific intervals, and other signals calling for action at irregular and occasional moments. With some individual differences of detail, the splitting up proceeded regularly from margin to centre. The merely occasional stimuli were the first to break away from the rest. There was a phase during which they were met by delayed, and often hurried, response. At length they were very frequently indeed ignored, to use psychological language they were 'forgotten', and there was a definite and, as it might be called, 'stupid' lapse of action. The machine control was, for example, constructed as if it were maintained by an internal combustion engine. A device indicated the amount of petrol available at any time during a trial and replenishment was required irregularly. At first the petrol gauge was never neglected. As time went on it became common for the operator to put off replenishment, and eventually to

snatch hurriedly at the control lever. With increasing fatigue, over and over again, the petrol signal was ignored until the machine stopped and the experiment reached a temporary inglorious end.

I must make a brief comment on the initial guess about additional, or distracting cues. When the proprioceptive impulses which would normally be set up by movement of the machine were given to the alert operator, in by far the greater number of cases—75 %—they were at once interpreted in terms of the machine. There was first a deterioration in most aspects of the performance: but this was temporary only. Before long the impulses fell into place as additional and helpful stimuli, and the performance improved significantly. If, however, they were given to the tired operator, most of these—70 %—treated them as unwanted and disagreeable stimuli of bodily origin. There was now a temporary improvement, followed, more quickly as the operator was more tired, by marked deterioration. They were, in fact, not now additional cues for action, but merely signs of growing weariness. Their first effect was to rouse the operator to greater effort, but this could not be maintained, and then they became only an obsessive awareness of bodily discomfort, and impeded the performance.

It has been shown that the normal course of the experiment demonstrated a regular and progressive tendency on the part of the subject to lower his standard. This could be checked at any stage by introducing new instructions of a difficult type, the result of which was that while the operator was in fact doing just the same as before, it appeared to him that the whole task was made harder. When the alert subject had these instructions, he quickly improved and the improvement was maintained for a long time. If the tired operator were set the apparently more difficult task, he too improved, but only temporarily. Soon he slipped back to a lower level than before. In other words he could, within the limits of fatigue set by the experiment, still carry out the local actions of control as well as or better than ever; but he could not maintain the organized, co-ordinated and timed responses for more than a short period.

Side by side with all these changes, recorded quantitatively and objectively, went some striking subjective phenomena. Three of these are of primary importance.

First the operator's reports of what actually happened during the experiment became less and less reliable as his fatigue increased. His falsifications were of all kinds. Events which actually happened failed to be observed. Changes that never happened at all—loud noises, uncontrollable variations of the instruments—were reported. Levers and switches were said to have become 'sticky', or 'heavy', or 'sloppy', or 'ineffective', though there was no change from beginning to end of the experiment. When the operator realized that he was making mistakes, he regularly maintained his belief that he himself was doing just as well as ever by blaming the experimenter or the machine.

Secondly, as already indicated, awareness of physical discomfort increased enormously. Appliances worn by the operator were charged with being very heavy, or uncomfortably tight. The temperature was said to be too hot, or too cold.

Postural discomfort came into the front of the picture, and again and again bodily cramp was reported. All the proprioceptive signs became more pressing and at the same time their interpretation became less accurate.

Thirdly, this experiment demonstrated conclusively that the everyday observation which associates growing irritability with increasing fatigue of the central nervous system is correct, and that all those masses of experiments which, concentrating upon some simple, repeated mental task, have failed to reveal this are wrong. When the operator began, absorbed in his task, he was usually silent. As he went on, sighs and shufflings emerged from the machine. Then mild expletives took the place of sighs. By the end of the experiment, which lasted for two hours or more, most operators kept up a flow of the most violent language they knew. And all the time their handling of the controls became more and more rough, so that they were doing more work, and not less, as the task went on. It was at this stage that the tendency to project all errors on to the experimenter or the machine reached its height.

Here, then, in broad outline, is the picture of fatigue following highly skilled work. It was necessary to draw this picture before anything certain or convincing can be said about the fatigue specific to prolonged exposure to loud noise or to vibration, to extremes of temperature, to lack of oxygen, to drugs, or to the many other special circumstances which are reasonably suspected to make the skilled man tired.

I must now try to bring together all these points and see what they tell us about how the central nervous system reacts to a complex environment when a man is growing weary. Central nervous fatigue cannot be evaluated in terms of a lessening of total effort, or, except in extreme stages perhaps, by any determination of the efficiency of the local reaction mechanisms. Its signs are indirect ones and are concerned with how the task is done, not with how much of it is done. As everybody knows, it is possible to take any mode of sensorial reaction—visual, auditory, tactual, thermal, proprioceptive—and find out the minimal value of effective stimulus necessary to evoke it and its changes. These are the sensory thresholds. As the senses and their modes of operation combine, under the guidance of central control, a new threshold emerges, a threshold of range or of limits, depending not upon the sensitivity of the special organs, but upon the place of the response in the task which is the combined expression of them all. This threshold of range appears as a standard, developed by the brain and used by it at every moment of every skilled performance. Yet it need not be formulated and the operator may remain as unaware of it as are the eye or the ear of their thresholds. When the operator tires, his standard drops, or, in other words, the threshold of effective range widens and becomes, in fact, the best of all single measures of his fatigue.

We can, perhaps, now begin to see how it is that timing is of tremendous importance as a sign of fatigue of this type. Time is not of great significance in relation to the single reaction. It is true that this has its history; it begins, rises to its maximum and fades away. But time does not matter much unless the single

reaction must be followed by another and the two together have to achieve a practical fit. As tasks become more important than single actions, and central control dominates local performance, the temporal relations of activities acquire a practical importance equal to that of the facility of the constituent responses themselves. Time is a discovery, running side by side with that of the co-ordination of actions and achieved by the same central control. If the threshold of effective range for any leading constituent of skill is raised, other constituents will be pushed further off from it in the order of successive performance, and then these other constituents will perhaps be hurried or forgotten or the whole performance will take longer. So it is the timing that goes wrong more easily than the efficiency of the local reactions. The rhythm of sequence of the activity is lost or broken, and performance becomes irregular, a story of spurts and delays.

All skill exhibits a number of constituent reactions which fit together. Some skills, when they are unrolled, are normally regarded as complete in themselves, forming, so to speak, a single unit of performance. Others involve critical points that are something like changes of direction in a journey, and then the changes that occur in the skill at these critical points come to be treated as the finish of one unit in the complete accomplishment or the beginning of another. Now the standard, involving the threshold of range, operates in every unit of skill; but the combination of several units into a total accomplishment represents a still more advanced achievement.

When, many years ago, Hughlings Jackson introduced the notion of 'levels' of function of the central nervous system, he pointed out that certain very high levels seemed to require a process which he called 'formulation'. It appears to be fair to hold that the successful maintenance of a fluent spacing of units of skill in a complex accomplishment is at this high level, so that what is an unwitting tribute to time in the unit becomes a conscious and controlling schedule of time in the combination of units. When anybody is solely concerned with the single skill, however complex it may be, without regard to what precedes or follows it, time means nothing to him, however much it may signify to anybody who may be watching his behaviour. But he can fit one unit to another in fully adaptive performance only by the aid of a formulated time schedule. This is a late achievement of the brain and very unstable. When a man gets tired the skill units dissociate, the time schedule, if, as in the experiment, one is provided, slips back into the time standard, and if, during its progress, or when the unit is finished, he is required to estimate its length in a formulated time order, he goes wildly wrong.

It is very tempting to believe that the fitting of units of skill into combined action first becomes urgent in connexion with concerted behaviour, where what one man does in a complex skilled production must harmonize with what another does. Then, while the time standard is an individual phenomenon, the time schedule is a social discovery, as many have held to be the case. But of this there is, of course, no proof.

The splitting of the stimulus field, and the order of its disintegration, are merely the environmental sides of all this. It is a character of all skill stimuli to be

patterned, with certain central features dominant. The threshold range, rising for the central signals, gives less time for all the others, and those in the extreme margin get crowded out, while the others have to be dealt with in a hurry.

The fact that sensations of bodily origin become both more insistent and less attached to the skill is exactly what would be expected if this general picture is correct. There are two important points.

The proprioceptive impulses concerned are, in this case, additional stimuli, not, in fact, directly required by the skill, but set up in the course of the performance and capable of being used to improve the performance. When the stimuli and the skill split up, under the influence of increasing fatigue, the proprioceptive signals, like all the other stimuli, get accepted less and less in their relations to all the others, and more and more as individual occurrences. They then move in the direction of becoming veritable distractions, for they are not to be found among the original stimuli that set the skill going. If this were all that happened they would tend to become more and more marginal and, in accordance with the general rule, they might be expected to be increasingly ignored. But it is not all that happens. All the time the general drift in the operator is towards a less closely organized and effective central control. He is moving back towards a stage of behaviour which has many of the characters of an early phase in the history of increasing neuro-muscular guidance. Evidence from many other sources shows that in this early stage sensations of bodily origin play a large part in directing behaviour. With the increase of weariness they get back their character of being pressing, interesting, very inadequately understood, less tied up with specific performance and more individual occurrences, attractive on their own account. They are not, therefore, easily ignored, but their function is persistently limited to their occurrence.

It seems possible to make a generalization about distractions and fatigue, though this must await more exact experimental confirmation. If distractions, or additional stimuli, can be woven into the general stimulus context the alert nervous system will use them, and after a very brief initial disturbance, the skill will improve. With the onset of fatigue they fall apart from the performance once more, but if they represent important general factors in the acquisition of special forms of muscle control they cannot be ignored. Their threat is met with the characteristic increased effort of the early phases of central fatigue, but this effort will not be maintained. They then become hampering, obsessive and disagreeable.

At first sight the subjective symptoms of the fatigue characteristic of long continued skill all seem highly paradoxical. The widening of threshold range, the increasing inaccuracy of timed response, the dissociation of performance and of stimulus units—all these proceed or may proceed unwittingly. Yet the operator becomes more and more unreliable about the facts of his situation, he is more worried and possessed by physical discomfort, and his irritability grows. He is at once more optimistic about his performance and pessimistic about his state. He asserts that he is doing well, and at the same time he blames something or somebody

else for making him do badly. The explanation perhaps is simple. He does not know *how* his bodily mechanisms are operating to achieve the required skill. Hardly anybody ever does, for skill does not depend on that kind of knowledge. He does know or suspect, however, that the task is not proceeding in quite the same way as at first and so he must have his excuses.

This is the beginning of the story of fatigue following highly skilled work. It is only the beginning and only the broad outlines are as yet available. If the details are to be filled in there must be more and more experiments. At least we know how to shape these experiments. We know that we must look less at what the fatigued man does, and more at how he does it. We must adopt as our criteria, not some measure of total effort, and not any necessary deterioration of specific and repeated reaction, but the widening of the threshold of range, the loss of accurate timing, the splitting up of stimulus and response patterns.

Perhaps it is not unfair to suggest also that this experiment has applications far beyond the limits of its own immediate purpose. For generations investigators have sought a fuller understanding of the functions of central nervous control. Frequently they have upset such control by fatigue, by drugs, by cutting out bits of brain structure, and then have looked to see what is lost in the way of behaviour. Perhaps this is wrong. Perhaps we should try first and foremost, not to see what is lost or even what is left, but how that which is held is maintained.

The experimental investigations upon a study of which this lecture is mainly based was carried out at the request and with the support of the Flying Personnel Research Committee.

The effects of cold and colchicine on mitosis in the newt

BY H. N. BARBER AND H. G. CALLAN

John Innes Horticultural Institution, Merton, London

(Communicated by C. D. Darlington, F.R.S.)

—Received 5 March 1942—Amended 27 August 1942)

[Plates 9–11.]

The application of cold or colchicine to dividing epidermal cells of the newt larva produces by similar means an arrest of mitosis at metaphase.

The cells respond to both agents by complete suppression or by abnormal development of the spindle. The various types of abnormal spindle which are produced can be described in terms of the complete or incomplete inactivation of centromeres, centrosomes or both. Spermatogonia and spermatocytes, unlike the epidermal cells, invariably show complete spindle suppression during cold treatment.

The more obvious differences between the effects of the two agents are due (a) to the greater severity of spindle suppression produced by colchicine, and (b) to the slowing up of all cell processes produced by cold. Thus while both treatments lead to the production of relapsed tetraploid cells, these occur more commonly after colchicine treatment.

The untreated epidermal cell swells up during anaphase. This swelling is exaggerated in cells held at metaphase under the action of cold or colchicine.

The inactivation effects can be described in terms of a tactoid theory, as due to surface changes in centrosomes or centromeres which prevent them orientating or breaking down spindle molecules.

1. INTRODUCTION

The progress of cell division may be modified or suppressed by a great variety of agents. Subnormal temperature is known to interfere with the mitotic spindle in plant cells (e.g. Blakeslee & Cartledge 1927, on *Datura*; Darlington & La Cour 1940, on *Trillium*, etc.). Various drugs, including colchicine, have comparable effects.

The action of colchicine was first noticed by Dixon (1905) in leucocytes. Dixon interpreted the increase he found in the frequency of metaphases as due to the action of the drug in stimulating mitosis. Recently extensive and detailed studies have been carried out by various workers, notably Lits (1934), Dustin (1934), Brues & Cohen (1936) and Ludford (1936). Brues & Cohen were able to show that there was a precise concentration of colchicine in the presence of which regenerating liver cells of the mouse would pass into mitotic activity at the normal rate but would not proceed beyond metaphase; this stage therefore accumulated throughout the tissue. Below this concentration the rate of entry of cells into mitosis was normal, but some of the cells which reached metaphase were able to continue into anaphase and telophase; above this concentration there was a depressant action on the cells about to enter mitosis. Ludford determined the effective concentrations of various drugs necessary to inhibit anaphase in tissue cultures, and demon-

strated the quite remarkable specificity of action of colchicine at great dilutions (up to 1 part in 100,000,000).

The effects of colchicine on plant tissues have recently become widely known owing to their utilization in producing polyploids (Blakeslee & Avery 1937; cf. Darlington & La Cour 1942). This action has been shown to have the same type of cell effect as that described for animals. Thus with some variation in threshold between different species, it leads to complete suppression of the spindle. The chromosomes reach metaphase or super-metaphase spiralization and lie scattered about the cell. After some delay the centromeres divide, but no separation of the sister chromatids into two independently organized nuclei takes place. A tetraploid cell is the result (Levan 1938). The effects of cold on plant mitosis are similar but again with differences of threshold (Darlington & La Cour 1940).

In general Ludford's work indicates that the action of colchicine on chromosome movement takes effect in a similar way in animal cells. It seemed likely, however, that in view of the difference in organization of the animal spindle, the similarity in action of cold and colchicine which held good in plants might break down in animals.

We have therefore carried out comparative tests with an animal chosen for the accuracy with which we have found it possible to record and describe the movements of the chromosomes, namely, the newt (*Triton vulgaris*, $2n = 24$). In the following account we shall deal first with the observed and qualitative effects of treatment together with their quantitative recording and the immediate developmental inferences which have to be drawn. Secondly, we shall deal with the more indirect inferences about cell mechanics which are revealed by the type of breakdown the spindle undergoes.

The experiments were carried out on half-grown larvae. In the cold-treatment experiments, larvae were kept in water at a temperature of 3° C. Lower temperatures completely suppressed mitosis. Colchicine treatment consisted in the immersion of larvae in a 1 % solution at room temperature (18–20° C).

Fixations of amputated tail tips were made in acetic alcohol after various periods of treatment. The tails were stained in Belling's aceto-carmin, or by means of the Feulgen technique, and pressed out for examination of the epithelial cells. The same methods were used with testes.

2. THE ARREST OF MITOSIS AT METAPHASE

The most striking effect of cold and of colchicine is the arrest of mitosis when the chromosomes have reached metaphase spiralization. This arrest leads to an accumulation of cells at this stage. In figures 1 and 2 counts of the numbers of cells at the various mitotic stages in a standard area of tail (the tail fin epidermis behind the anus in half-grown larvae of a standard size) are plotted against the time of treatment.

Cold, it will be seen (figure 1, table 1) has an immediate and general depressant

effect on the initiation of mitosis. The number of prophases is reduced by 75 % within 12 hr., and remains more or less constant at this low level for over 4 days. But the most significant effect is that on the anaphases. After 12 hr. they disappear completely. The number of metaphases, on the other hand, increases from 33 in the controls to 86 in a period of 8 days. Evidently the cells coming into mitosis are held at metaphase. The metaphases are of various types, which we shall consider later.

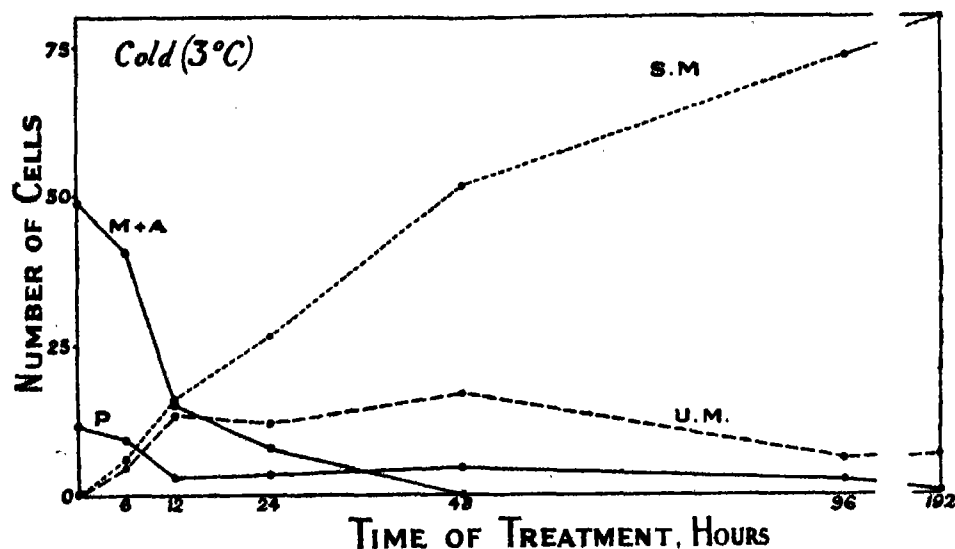


FIGURE 1. Graph showing frequency of cell types in the tail epidermis of the newt plotted against duration of cold treatment (3° C). The data are from table 1. The four lines give number of prophases (P.), metaphases with bipolar spindles and anaphases (M. + A.), and two categories of arrested metaphase, the *star* (S.M.) and *unorientated* metaphases (U.M.).

TABLE 1. FREQUENCY OF DIFFERENT TYPES OF CELL AFTER COLD TREATMENT (3° C). THE FIGURES ARE THE MEANS OF COUNTS OF THE CELLS IN THE POST-ANAL TAIL FINS OF THREE LARVAE

time of treatment hr.	prophase	bipolar metaphase	star metaphase	unorientated metaphase	total metaphase	anaphase
0	11.7	33.3	—	—	33.3	15.3
6	9.3	31.0	6.0	4.3	41.3	9.3
12	3.0	12.3	16.0	13.3	41.7	3.0
24	3.7	7.3	26.3	12.0	49.0	—
48	4.3	—	51.7	17.0	68.7	—
96	2.3	—	73.3	6.7	80.0	—
192	0.7	—	80.0	6.3	86.3	—

The initial effects of colchicine are similar (figure 2, table 2). They differ from those of cold in that colchicine does not at once reduce the number of cells coming into prophase. There is, however, an immediate reduction in the number of anaphases, and the few that continue frequently fail to separate the two daughter

nuclei completely (figure 21, plate 10). There is an even greater accumulation of cells with metaphase chromosomes. This means that the chromosomes are arrested at metaphase as they are during cold treatment, but that prophase is not slowed down immediately by the drug as it is by the low temperature.

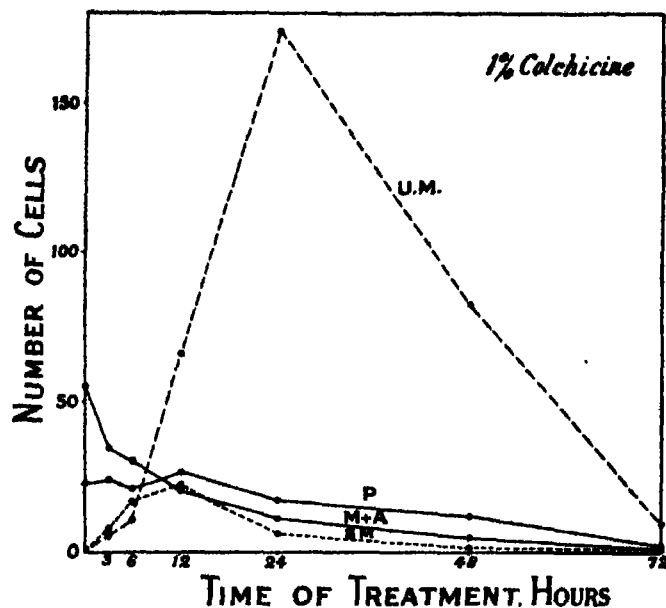


FIGURE 2. Graph showing changes in cell frequencies plotted against duration of colchicine treatment. The data are from table 2.

TABLE 2. FREQUENCY OF DIFFERENT TYPES OF CELL AFTER COLCHICINE TREATMENT (MEANS OF COUNTS FROM 3 LARVAE)

time of treatment hr.	prophase	bipolar metaphase	star metaphase	unorientated metaphase	total metaphase	anaphase
0	22.3	25.0	—	—	25.0	30.7
3	24.0	15.7	7.7	6.3	29.7	20.0
6	20.3	15.0	16.3	10.7	42.0	15.7
12	27.0	12.3	20.7	66.3	99.3	8.3
24	17.7	5.0	6.7	175.3	186.0	6.7
48	12.0	0.3	1.7	83.3	85.3	4.3
72	2.3	—	—	9.7	9.7	1.0

3. THE ABNORMAL CELL TYPES

Cold and colchicine have no visible effect on the form or manner of prophase (figure 3). But among the arrested metaphases the following types of abnormal chromosome development can be distinguished:

(a) The *Star* metaphase (figure 6; figure 11, plate 9). The undivided centromeres come together at one focus with the arms of the chromosomes radiating outwards in all directions. Odd chromosomes are frequently left out (figures 13, 14, plate 9).

In a small proportion of such cells two or three centromere-foci have been seen to give a pseudo-anaphase appearance (figure 15, plate 9).

(b) The *Distorted-Star* metaphase (figures 16, 17, plate 10). This type is similar to the star except that it is not radially symmetrical. The centromeres come

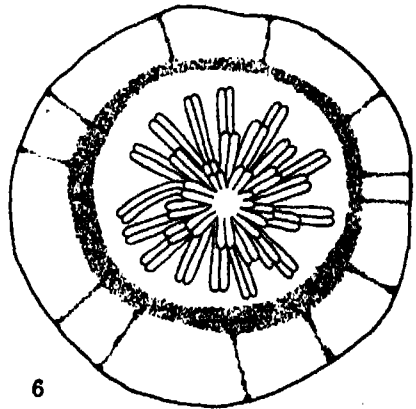
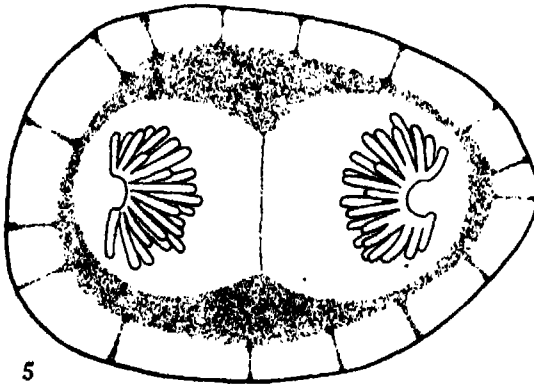
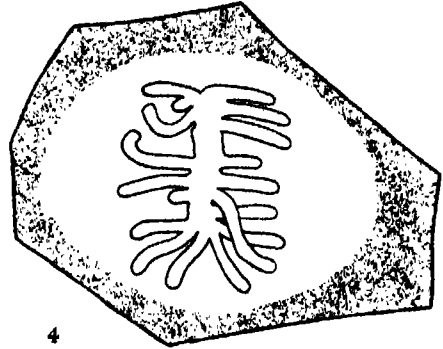
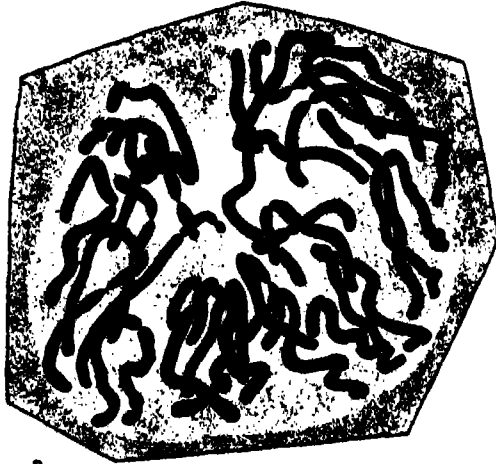


FIGURE 3. Normal prophase. Note disposition of centromeres as derived from the preceding telophase, and absence of cytoplasmic vacuoles. Aceto-carmin. $\times 1600$.

FIGURE 4. Normal metaphase. Note absence of cytoplasmic vacuoles. Iodine-alcohol. $\times 1200$.

FIGURE 5. Normal anaphase. Note vacuolization of cytoplasm outside spindle area. Iodine-alcohol. $\times 1200$.

FIGURE 6. *Star* metaphase after 5 days at 3°C . Note vacuolation of cytoplasm outside the spindle area. Iodine-alcohol. $\times 1200$.

together at one focus, but at one side there is a gap around which the chromosome arms are stretched. This leads in the extreme condition to a flattened star pushed against one side of the cell.

(c) The *Exploded* metaphase (figure 18, plate 10). The centromeres remain completely unorientated, but the chromosomes are pushed bodily to the boundary of

the cell, many becoming stretched. The centromeres themselves may or may not be stretched, indicating that they play no specific part in the movement. In the final condition the chromosomes may become segregated in several groups, but no separation of sister chromatids into two different nuclei takes place (figure 19, plate 10). The products of this segregation are unbalanced and ultimately degenerate.

(d) The *Prophase* metaphase (figure 9, plate 9). The fully contracted metaphase chromosomes retain the prophase orientation, with the centromeres directed to one side of the nucleus. According to the compactness of the resting nucleus from which they have come, so the chromosomes lie close together or far apart. This type corresponds to the transitory 'prometaphase' in untreated cells, but owing to the failure of spindle development it is a fairly stable condition under treatment.

(e) The *Ball* metaphase (figure 10, plate 9). The chromosomes form a closely tangled ball in the middle of the cell.

Types *a* and *b* are grouped as Stars in the graphs, types *c*, *d* and *e* as unorientated metaphases.

The types of metaphase selected for description are of course arbitrary and show a regular intergradation, such as they no doubt sometimes undergo in development. Thus the prophase-metaphase seems to pass direct into the ball. In this cell type it is often possible to distinguish traces of prophase orientation of the centromeres. Similarly the star, which in its typical form shows the arms radiating straight out, may also pass by relaxation of the orientation into the ball (figure 8).

All these types of cell occur after treatment either with cold or with colchicine. But their proportions differ. The commonest type in the cold is the star (over 90 %). In colchicine on the other hand, the ball is the commonest (75 %), and stars as a rule appear only in the early fixations (up to 24 hr.). The other types are rarer, the exploded type occurring sporadically throughout the period of treatment.

Spermatogonia and spermatocytes were similarly subjected to cold treatment. They also are arrested at metaphase. Unlike epidermal cells, however, all the arrested metaphases are of the prophase-metaphase type.

Both cold and colchicine treatments frequently seem to disturb the gross timing relations of the processes responsible for cell division. This appears in two ways. First both star and exploded metaphases may be formed by chromosomes whose spiralization has reached only the mid-prophase state (figure 12, plate 9). Possibly the precocious breakdown of the nuclear membrane, whose equilibrium during prophase is inherently unstable, is the conditioning factor.

The second timing irregularity is shown by what we may call *swelling*. Previous to anaphase in the untreated epithelial cells, the cell outline is polygonal (figures 3, 4). During anaphase, however, the dividing cell swells by the intake of liquid and forms an even, rounded outline at the expense of the neighbouring cells. This liquid forms vacuoles in the cytoplasm immediately inside the cell boundary and outside the spindle (figure 5). The vacuolation of the anaphase cytoplasm may be related to 'bubbling' at anaphase described by Strangeways (1922) in chick fibroblasts and by Barber (1939) in *Tradescantia* hair cells.

The various types of cell containing arrested metaphases invariably develop this swelling to excess (figure 6). Their cytoplasm is in fact in a condition of prolonged and thereby exaggerated anaphase although the chromosomes are still in metaphase. The intake of water may be due to an increase of osmotic pressure inside the cell membrane; this increase may be ascribed to a repartitioning of the bound and free water during chromosome spiralization and spindle growth, resulting in a decrease of free water and hence an increase in solute concentration. Guinochet (1940) has demonstrated that under the action of colchicine the osmotic pressure of the cell sap of wheat seedlings increases from 7.1 to 11.1 atmospheres. Normally the ensuing telophase, with the breakdown of the spindle and the reconstitution of resting nuclei, will liberate the cytoplasm from this stress. The arrested metaphase cell, on the other hand, must remain abnormally swollen until its chromosomes lapse into a resting stage or until disintegration takes place. In consequence cold and, even more, colchicine seem to disturb the gross water relations of the larvae. After heavy treatment the whole animal often swells enormously. We have as it were induced intracellular dropsy.

Under continued treatment anaphase is frustrated in all types of arrested cells. In colchicine the arrest is followed after a few hours by a lapse of the chromosomes into a normal resting nucleus or by degeneration into pycnotic masses. The sequence of events is similar to that described by Levan (1938) in *Allium*. The chromatid attraction lapses and division of the centromere takes place, but the two chromatids remain parallel. Either a single resting nucleus, which will be tetraploid, is produced, or several unbalanced nuclei which will degenerate.

In the cold, on the other hand, the arrested metaphases persist for several days and the return to the resting stage so as to produce a tetraploid cell (figure 20, plate 10) is much rarer, or slower, than with colchicine. Evidently the cold holds up not only the spindle but all other cell processes with it. In particular, low temperature has an effect on the processes of despiralization and loss of attached nucleic acid by the chromosomes: many nuclei in metaphase or anaphase commence the telophase processes, only to degenerate before telophase is completed. Degenerating nuclei are characterized by their heavy staining and many-pointed periphery, associated with contraction.

4. RECOVERY

When larvae are taken from the cold and placed in water at room temperature (18–20° C), the arrested cells recover rapidly. After 1–2 hr. anaphases are frequent. Many of them are multipolar (figure 22, plate 11). Some have incompact spindles such as occur naturally in certain plants (figure 23, plate 11) (cf. Darlington & Thomas 1937, and Clarke 1940). In some recovered anaphases, after colchicine as well as after cold, the chromatids stick, either terminally or elsewhere. Occasionally, however, sticking has also been observed in the controls (figure 7; figures 25, 27, plate 11).

The sticking seems to lead to a structural change. For example acentric fragments (figure 26, plate 11) sometimes accompanied by a bridge, have been observed. Darlington & La Cour (1940) have shown a similar increase in the frequency of structural change after cold treatment in *Trillium*. Here the increase was correlated with the nucleic acid starvation of the heterochromatic regions of the chromosomes at low temperatures. Similar heterochromatic regions are revealed in *Triton* under these conditions (Callan 1942), and the structural changes may thus be caused by the same reproductive error as in *Trillium*.

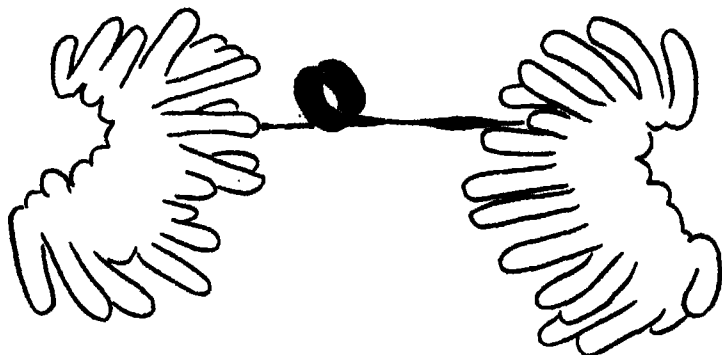


FIGURE 7. Intercalary and terminal (?) sticking at anaphase.
No treatment. Aceto-carmin. $\times 1600$.

Recovery after colchicine treatment is much slower than after cold treatment. The drug remains in the cells; the temperature can be changed at once. In consequence the larvae rarely recover completely from the colchicine treatment; nearly all of them die. Among the survivors only a few divisions have been seen. Most of these are diploid, but one approximately tetraploid cell with evidence of secondary pairing was seen in a fixation 5 days after the end of the treatment. Later fixations showed only diploid divisions.

5. THE ORIGIN AND STRUCTURE OF THE SPINDLE

Numerous theories have been put forward to account for the origin and properties of the spindle (cf. Gray 1931). Most of these speculations were premature attempts to explain the spindle in terms of physico-chemical concepts.

The earlier investigators (Boveri, Wilson, etc.) showed that in animal cells spindles are formed in relation to extranuclear bodies, the centrosomes. In plants, except in some of the lower groups (Algae, Fungi) centrosomes have not been recognized. However, as Darlington & Thomas (1937) have shown in a grass, corresponding pole determinants must exist in the higher plants, but they are more diffuse than in animal cells.

In many animal cells the centrosomes have continuity from one division to the next. It therefore seemed justifiable to speak of them as cell organs, particularly

in those cases where, as basal granules, they are associated with cilia or flagella. However, the evidence from experiment runs counter to this view. Morgan (1896) and McClendon (1908), by exposing echinoderm eggs to hypertonic sea water, have induced subsidiary cytasters associated with polar granules which seem to have arisen *de novo* in the cell. The evidence indicated that this agent had not acted by stimulating the division of preexisting centrosomes, but rather by altering the surface properties of other granules already distributed in the cell. Similarly the surface of the cell can take over centrosome functions, as in the 'monopolar' spindles of *Sciara* (Metz 1933).

The centrosome, or its equivalent, is however not the only spindle organizer in the cell. Normal spindles, as was implicit in Bölaß's 'Stemmkörper' theory, are formed by the co-operation of centrosomes and the intrachromosomal centromeres. In different organisms one or other of these organizers may fail to operate properly. Thus at meiosis in hybrids with little chromosome pairing the centrosomes may organize a spindle to which the univalent chromosomes fail to attach themselves. These spindles are characterized by being narrow and too long for the cell. As they grow they become contorted (e.g. Beadle 1933, on a mutant *Zea*; Dobzhansky 1934, on *Drosophila* hybrids; and Darlington & Thomas 1937, on a grass). The growth of this purely centrosome spindle may push the chromosomes about the cell just as acentric fragments are pushed about by the growth of a normal spindle. The centromeres remain inactive.

In other organisms the centrosomes may fail. Thus in the grass hybrid (Darlington & Thomas 1937), in a mutant *Zea* (Clarke 1940), or in some coccids as a normal process (Schrader 1932), purely centromere spindles are formed at meiosis. These spindles are characterized by being made up of several discrete spindles, one per bivalent, arranged more or less parallel, but often diverging at the ends. Each subsidiary spindle has the structure of a centrosome spindle except that the fibrils diverge from the centromere. The centromere is behaving like a centrosome. When anaphase occurs, several micronuclei may form at the ends of the cell instead of a single nucleus. Evidently the centrosome is necessary in most cells to bring the spindle to a focus at the poles.

We have seen that under the influence of cold or colchicine several abnormal types of chromosome movement and orientation take place in the epidermal cells of the newt. These various types can readily be interpreted in terms of varying degrees of centromere and centrosome activity (figure 8).

In the most extreme types, viz. the ball and prophase metaphase, both centrosome and centromere are inactivated. No directed chromosome movement takes place, and the movements which do occur are internal. They result from the coiling and uncoiling of the chromosomes and the lapse of chromatid attraction.

In the exploded metaphase the chromosomes are pushed bodily about the cell, probably owing to the formation of a purely centrosome spindle. The centromeres are inactive, and the chromosomes thus passive in their reactions to this spindle. The centrosome spindle may be a monaster, when the chromosomes are pushed

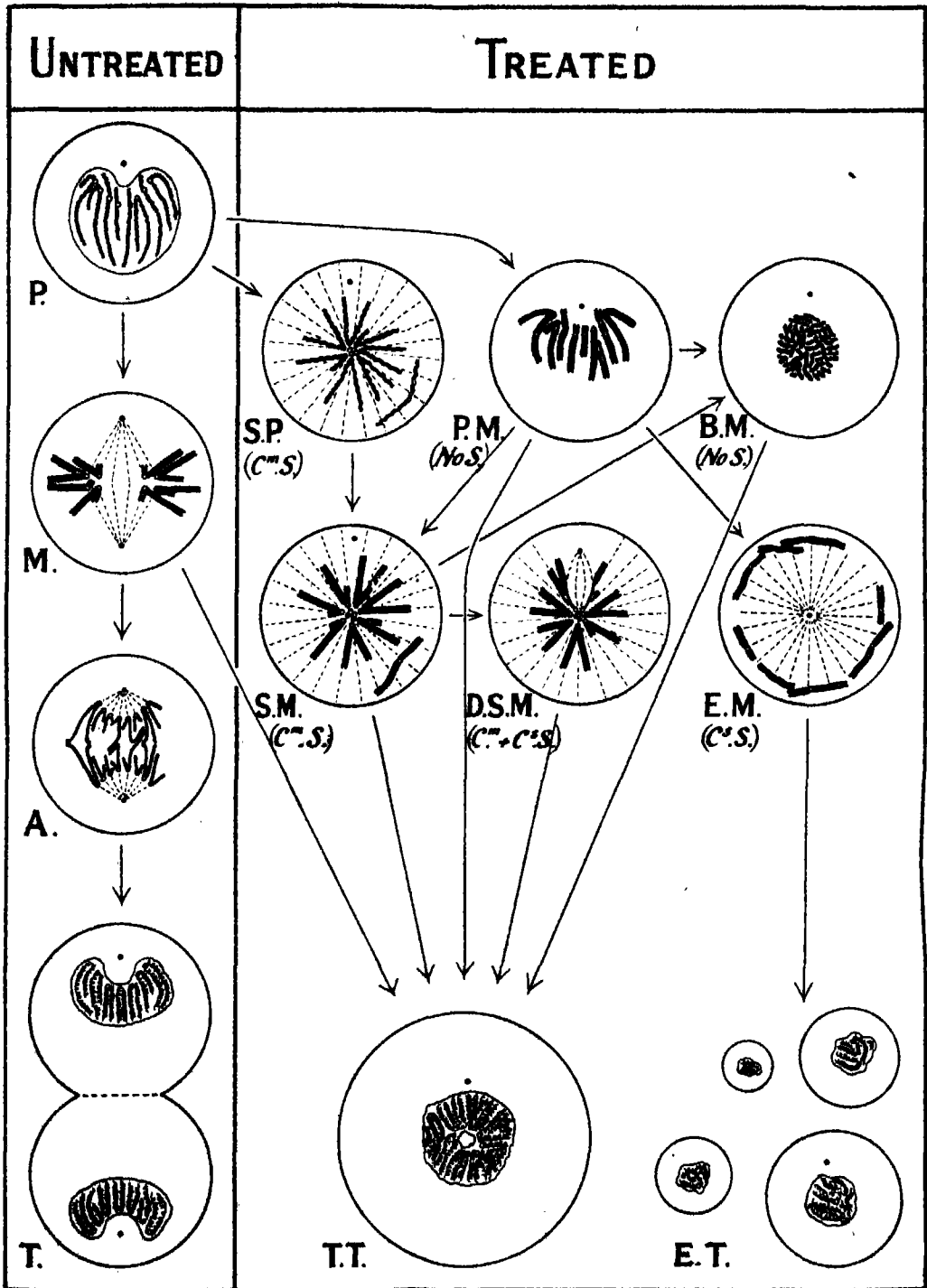


FIGURE 8. Diagram showing the relationships of types of spindle produced after cold and colchicine treatment of the newt larva.

P. = prophase, *M.* = metaphase, *A.* = anaphase, *T.* = telophase, *S.P.* = star prophase, *P.M.* = prophase metaphase, *B.M.* = ball metaphase, *S.M.* = star metaphase, *D.S.M.* = distorted star metaphase, *E.M.* = exploded metaphase, *T.T.* = tetraploid telophase, *E.T.* = exploded telophase, $C^m.S$ = centromere spindle, $C'.S$ = centrosome spindle.

to the edge of the cell, or it may have peripheral poles, when the chromosomes are separated into clumps which simulate anaphase.

At first sight the star metaphase might also be interpreted as a monopolar centrosome spindle on which the centromeres have orientated. However, Bélař (1933) and Fankhauser (1934 *a, b*) have shown that the chromosomes become orientated on such a spindle at a distance from the centre. The metaphase plate lies on the surface of a sphere. In the star, on the other hand, the centromeres come to lie as close to the focus as possible, and we conclude that they are themselves organizing spindles which close to form a monaster. Possibly the proximity of the centromeres to one another in the prophase nucleus accounts for the normal formation of a single common spindle.

The distorted star probably results from an abnormal co-operation of the centromeres with the centrosome. In this type of arrested metaphase a dipolar spindle is present, one pole of which is formed by the centromeres, the other by the centrosome. The central spindle grows at the expense of the centromere monaster, thus forcing the chromosomes against the cell boundary, where they form a star flattened in one plane.

So far we have discussed the spindle without venturing any speculations as to its physical or chemical nature.

Bernal (1940) has put forward a new theory of the spindle based on the behaviour of liquid crystals made up of protein fibres. He assumes the cyclical formation and dissolution of 'tactoids' in the cell. These spindle-shaped bodies are formed when long molecules reaching a threshold concentration come to a position of equilibrium in a medium of unorientated molecules. In brief the long molecules come to lie parallel to one another as a result of long range Langmuir-Levine forces. These attraction forces exist owing to the charges of each particle interacting on the ionic atmospheres of other particles. The existence of an orientation of long particles on a free surface produces a surface tension which has greater power to bend the surface perpendicular to the axis of orientation than along that axis. The length and diameter of the tactoid are a function of the length of its component fibre molecules. The longer the molecules, the longer are the tactoids of the same cross-sectional area.

Bernal further points out that an enclosure of a non-orientated liquid within a tactoid medium itself assumes spindle form. This is termed a negative tactoid.

Two lines of evidence favour the view that the mitotic spindle is a tactoid. Bélař (1929) exposed living spermatocytes of *Stenobothrus* to hypertonic liquids, and found that the resultant shrinkage of the spindle was greater transversely across the spindle than along its axis. This indicates that the spindle is made up of long particles orientated parallel to the main axis. Further, Schmidt (1937) has shown that living spindles are positively birefringent, which again indicates the same type of particle orientation.

Bernal suggests that at mitotic prophase a progressive chemical change occurs within the cell which results in the formation of long protein molecules. As prophase

proceeds the particles become longer and form parallel chains. Owing to certain surface characteristics of the centrosome granules, these chains come to lie perpendicular to the centrosome surface. Such a condition, the monaster, is inherently unstable, but with the division of the centrosome into two parts, the more stable spindle tactoid is formed. As the length of each individual particle increases, so the length of the spindle increases.

At this stage Bernal assumes that owing to a different surface characteristic of the centromeres, local dissolution of the long molecules of the spindle close to the centromeres occurs. Negative tactoids are thus formed, which pass into a position of equilibrium on the equatorial plate of the spindle, carrying the centromeres and chromosome arms with them. Now the centromeres divide: the negative tactoids increase in size, forcing the divided centromeres apart, this being the motive force of anaphase.

Finally, a retrogressive chemical change comes over the cell which involves the breakdown of the long molecules forming the spindle. The divided cell passes from telophase to a resting stage.

If we accept the general principles of Bernal's theory, certain new interpretations may be made in the light of our experiments with cold and colchicine and of other cytological observations. The surface properties of the centromere are not identical at all stages and in all cells. Probably in most cells, prior to metaphase, the surface of the centromere resembles that of the centrosome in that fibre molecules set perpendicularly to it. In plant cells with no centrosomes the wave of orientation of the fibre molecules originates at the centromeres. The centromere, however, unlike the centrosome, is not radially symmetrical, with the result that centromere spindles are not asters. The fibre molecules converge to two sides of the centromere only, forming a spindle shaped like two cones with their apices abutting. In cells possessing centrosomes, the centromeres co-operate to a greater or lesser extent in orienting the spindle molecules. At metaphase, however, the centrosomes have come to play the predominant role, perhaps owing to the changing surface conditions of the centromeres which are passing into the dissolution phase.

Cold and colchicine both act on the surface properties of centrosomes and centromeres, and the extent of their action varies according to the stage of the cell. (i) Cells which were at metaphase when the treatment started are prevented from passing to anaphase. This indicates that both agents prevent the change at the centromere surface and thus stop the local breakdown of the spindle and the formation of negative tactoids. (ii) Cells which enter mitosis during treatment may also show suppression of this centromere change, viz. star and distorted star metaphases. (iii) But the suppression may act earlier, by so affecting the surface of the centromere or centrosome or both that the fibre molecules do not set perpendicular to these surfaces. Thus in the exploded metaphase the centromeres, in the star the centrosomes, and finally in the ball and prophase metaphases both centromeres and centrosomes are inactivated.

REFERENCES

- Barber, H. N. 1939 *Chromosoma*, 1, 33.
 Beadle, G. W. 1933 *Cytologia, Tokyo*, 4, 269.
 Bělař, K. 1929 *Roux Arch. Entw. Mech. Organ.* 118, 359.
 Bělař, K. 1933 *Z. Zellforsch.* 17, 51.
 Bernal, J. D. 1940 *The cell and protoplasm*, p. 199.
 Blakeslee, A. F. & Avery, A. G. 1937 *J. Hered.* 28, 393.
 Blakeslee, A. F. & Cartledge, J. L. 1927 *Mem. Hort. Soc. N.Y.* 3, 305.
 Brues, A. M. & Cohen, A. 1936 *Biochem. J.* 30, 1363.
 Callan, H. G. 1942 *Proc. Roy. Soc. B*, 130, 324.
 Clarke, F. J. 1940 *Amer. J. Bot.* 27, 547.
 Darlington, C. D. & La Cour, L. F. 1940 *J. Genet.* 40, 185.
 Darlington, C. D. & La Cour, L. F. 1942 *The handling of chromosomes*. London.
 Darlington, C. D. & Thomas, P. T. 1937 *Ann. Bot., Lond., n.s.* 1, 747.
 Dixon, W. E. 1905 *A manual of pharmacology*. London.
 Dobzhansky, Th. 1934 *Z. Zellforsch.* 21, 169.
 Dustin, A. P. 1934 *Bull. Acad. Méd. Belg.* 14, 487.
 Fankhauser, G. 1934a *J. Exp. Zool.* 67, 349.
 Fankhauser, G. 1934b *J. Exp. Zool.* 68, 1.
 Gray, J. 1931 *Experimental cytology*. Cambridge.
 Guinochet, M. 1940 *C.R. Soc. Biol. Paris*, 210, 579.
 Levan, A. 1938 *Hereditas, Lund*, 24, 471.
 Lits, F. 1934 *C.R. Soc. Biol., Paris*, 115, 1421.
 Ludford, R. J. 1936 *Arch. Exp. Zellforsch.* 18, 411.
 McClendon, J. F. 1908 *Arch. EntMech. Org.* 26, 662.
 Metz, C. W. 1933 *Biol. Bull. Woods Hole*, 54, 333.
 Morgan, T. H. 1896 *Arch. EntMech. Org.* 3, 339.
 Schmidt, W. J. 1937 *Die Doppelbrechung von Karyoplasma, Zytoplasma und Metaplasma*. Berlin.
 Schrader, F. 1932 *Z. wiss. Zool.* 142, 620.
 Strangeways, T. S. P. 1922 *Proc. Roy. Soc. B*, 94, 137.

DESCRIPTION OF PLATES 9-11

Photographs all $\times 1750$ approx.

PLATE 9

- FIGURE 9. *Prophase metaphase*. 12 hr., 1 % colchicine. Feulgen.
 FIGURE 10. *Ball metaphase*. 24 hr., 1 % colchicine. Feulgen.
 FIGURE 11. *Star metaphase*. 4 days, 3° C. Aceto-carmin.
 FIGURE 12. *Star with prophase spiralization*. One chromosome has failed to orientate. 6 hr., 3° C. Feulgen.
 FIGURE 13. *Incomplete star with two foci*. 3 days, 3° C. Feulgen.
 FIGURE 14. *Incomplete star*. 6 hr., 1 % colchicine. Feulgen.
 FIGURE 15. *Simple and bifocal stars*. 12 hr., 1 % colchicine. Feulgen.



11



12



13

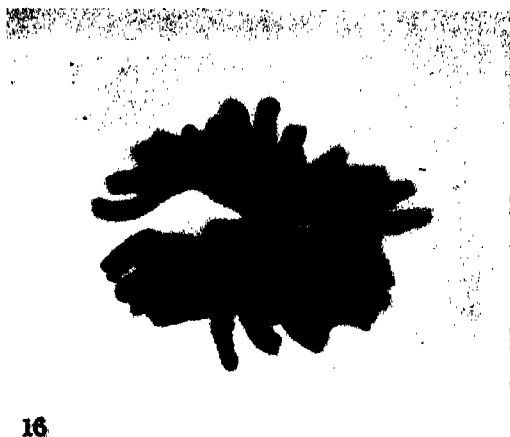


14



15







22



23



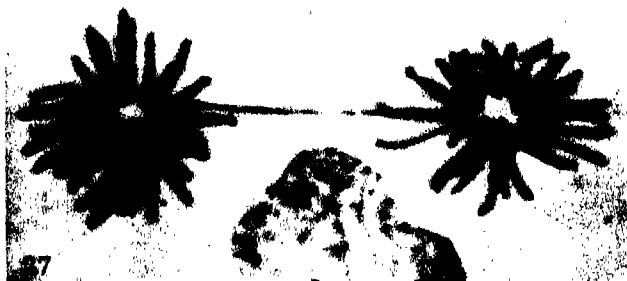
24



25



26



27

PLATE 10

FIGURE 16. Early stage *distorted star*. 6 hr., 1 % colchicine. Feulgen.

FIGURE 17. Later stage *distorted star*. 1 day, 3° C. Feulgen.

FIGURE 18. Early stage *exploded metaphase*; note stretching of the chromatids. 1 day, 1 % colchicine. Feulgen.

FIGURE 19. *Exploded nucleus*. Note deformation of surrounding resting nuclei and the pseudo-anaphase appearance of the groups of chromosomes. The centromere of one chromosome in the group on the left has divided. 2 days, 1 % colchicine. Feulgen.

FIGURE 20. Tetraploid resting nucleus resulting from relapse of an arrested metaphase. 2 days, 3° C. Feulgen.

FIGURE 21. Tetraploid telophase resulting from a partially arrested anaphase. 6 hr., 1 % colchicine. Feulgen.

PLATE 11

FIGURE 22. Quadripolar anaphase. 7 days, 3° C; 2 hr. 18° C. Aceto-carmin.

FIGURE 23. Anaphase with an incompact spindle. 4 days, 3° C; 2 hr., 18° C. Feulgen.

FIGURE 24. Unbalanced resting nuclei derived from a quadripolar anaphase. 4 days, 3° C; 3 hr., 18° C. Feulgen.

FIGURE 25. Intercalary sticking at anaphase. No treatment. Aceto-carmin.

FIGURE 26. Fragment at anaphase. 4 days, 3° C; 2 hr., 18° C. Aceto-carmin.

FIGURE 27. Terminal sticking at anaphase. No treatment. Feulgen.

The visceral nervous system of the earthworm:

I. Nerves controlling the tone of the alimentary canal

By FLIGHT-LIEUTENANT NORMAN MILLOTT, M.Sc., R.A.F.V.R.

*Lecturer in Zoology, The Victoria University, Manchester, formerly
Rouse Ball Student, Trinity College, Cambridge*

*From the Department of Zoology, Cambridge, and the Department
of Zoology, The University, Manchester*

(Communicated by H. Graham Cannon, F.R.S.—Received 15 January 1942)

The alimentary canal of *Lumbricus* exhibits autonomous rhythmic movements, probably co-ordinated by nerve elements situated in its walls.

The gut is subject to indirect control by extrinsic nerves which fall into two groups mutually antagonistic in their effects.

Nerves which augment the tone of the gut muscles leave the central nervous system by the middle and posterior nerves of each segment and join nerve elements (apparently arranged as a plexus) situated in the peritoneum of the body wall, from which nerves pass to enter the gut by the ventro-lateral regions of each septum. Nerves which diminish the tone of the gut leave the central nervous system by the anterior, middle and posterior nerves of each segment and join nerve elements (also probably in the form of a plexus)

situated in the muscular layer of the body wall, from which nerves arise to enter the gut via the dorso-lateral region of each septum.

The alimentary canal receives another nerve supply from the peripharyngeal commissures of the central nervous system by means of a number of fine plexiform nerves partially embedded in the pharyngeal musculature of either side (pharyngeal plexus). The anatomical relationships of these nerves are described.

The pharyngeal plexuses are shown to embody nerve elements which exert an indirect control over the gut, stimulation of these resulting in a fall in tone of all regions of the gut behind the pharynx.

The nerve supply of the gut is compared with that which is known in other annelids, and the mechanism of indirect control is compared with that existing in vertebrates.

INTRODUCTION

The alimentary canal of the earthworm, like that of any coelomate, has some characters in common with that of the vertebrates. In both, the gut lies freely in the body cavity and shows a relatively high degree of autonomy. In the vertebrates the gut possesses a diffuse nerve plexus immediately controlling its activities, and this network is in turn controlled by the central nervous system through two mainly antagonistic systems of nerves—the sympathetic and parasympathetic. In the gut of the earthworm, all earlier workers agree that there is a well-developed plexus of nerve cells within its walls, and to this extent at least it resembles the condition found in vertebrates. The question arises whether the resemblance extends further, and in particular, whether the gut of the earthworm is controlled by an extrinsic nerve supply from the central nervous system, as in the vertebrates. It is the object of this work to answer this question.

INNERVATION OF THE ALIMENTARY CANAL

There are few detailed accounts of the nerve supply of the gut of earthworms. The nerve plexus in the gut wall (enteric plexus) is described by Smirnow (1894), Dechant (1906), and Hess (1925). Several workers have described nerves passing to the gut from outside sources, but most of them are concerned only with nerves passing to the pharynx.

Quatrefages (1847) states that the pharynx is supplied by a plexus of nerves arising from the pharyngeal commissures of the central nervous system. Clark (1856) confirms this observation, and differentiates the plexus into inner and outer portions which supply, respectively, the muscles and mucosa, and the muscles and pharyngeal glands. Lankester (1865) gives a similar description.

These authors do not mention any further innervation of the alimentary canal, and both Lankester and Clark express the view that the pharyngeal plexus is a reflex centre, independent of the central nervous system but to some extent controlled by it, presiding over the functioning of the alimentary canal. Such a view agrees with those expressed by Dechant (1906), Hess (1925), and Wu (1939a): the two former regard the plexus as a special portion of the enteric plexus, while the latter claims to have shown that the pharyngeal plexus contains nerves which increase the tone of the crop and gizzard.

The innervation of other regions of the gut is described by Smallwood (1923), who claims to have discovered plexuses of nerves in the peritoneum and longitudinal muscles of the body wall from which plexuses of fine nerves pass to the gut via the septa. Plexuses of nerves investing the dilator muscles of the pharynx and the blood vessels supplying the gut are also stated to contribute nerves to the gut. Smallwood later confirmed these results by a different technique (1926). In addition, Wu (1939a) describes some experiments which are interpreted as showing that the crop and gizzard are supplied by inhibitor nerves from the underlying ventral nerve cord via the segmental nerves, and the adjoining septa and blood vessels.

In examining the significance of these two systems of innervation it will be shown that the gut, like that of vertebrates, is autonomous, and that therefore the control exerted by these systems is indirect. Secondly, it will be shown that this control is shared by both systems, and, coupled with the autonomous movements of the gut, appears sufficient to account for all the activities of the gut so far observed.

METHODS

Whenever possible the earthworm *Lumbricus terrestris* was used. Occasionally *Allolobophora longa* was used, but no difference was found between the two forms which has any significance in this account.

The behaviour of all regions of the gut except the pharynx was examined, and during experiments all exposed regions of the gut were bathed, or kept constantly moist, with Ringer's solution of the following percentage composition:

NaCl	0.65	CaCl ₂	0.012	NaH ₂ PO ₄	0.001
KCl	0.014	NaHCO ₃	0.02	Water	to 100

This fluid served excellently, despite the fact that its osmotic pressure is considerably higher than that of the body fluids of earthworm as estimated by Adolf (1927). To allay possible objection on these grounds, a number of experiments were repeated with the gut bathed in a fluid made by diluting the above with one-third of its volume of distilled water (this makes a solution with about the same osmotic pressure as that given by Adolf for the body fluids of earthworm). Such modification made no difference to the results of the experiments, and therefore in most cases the unmodified Ringer was used.

In some experiments the gut was excised and observed whilst in a bath of oxygenated Ringer, but in most cases the gut was exposed *in situ*, by opening up the appropriate region of the worm (unless otherwise stated in the mid-dorsal line), and kept moist with Ringer. When using the latter method precautions were taken. First, before starting an experiment all the coelomic fluid was washed away from the region of the gut to be examined, thus preventing the possible deleterious effects of coagulation. Secondly, the body wall of the region opened

up was stretched and pinned down very firmly. This was found to be sufficient to eliminate movements of the body wall which might otherwise have interfered with the recording of the movements of the gut (see p. 279).

The gut was attached (in the manner shown in figure 4) to a light isotonic lever writing on a slowly revolving smoked drum. Since the worm was stretched out in a horizontal plane, the movements directly recorded were those of the circular muscles of the gut, those of the longitudinal muscles modifying the record by altering the shape of the gut. The recorded responses of the muscles were checked by direct observation.

In vertebrates, valuable information concerning the peripheral nervous system can be obtained by the use of drugs. Analogous experiments were performed here, in which drugs having a well-defined action on the vertebrate nervous system were employed.

AUTONOMOUS MOVEMENTS OF THE GUT

A detailed account of the movements of excised portions of the gut is given by Wu (1939 a).

When any portion of the alimentary canal of a freshly opened worm is examined *in situ* it exhibits active peristaltic movements. These movements vary greatly in rate and amplitude in different individuals and in different regions of the gut. Thus the oesophagus, and in particular the crop, show most active movements, while the intestine is almost invariably the most sluggish. Frequently the movements are markedly irregular, but it is possible that such movements are abnormal and due to the experimental conditions. Active anti-peristalsis is often observed in the region of the calciferous glands.

The peristaltic movements persist, though much reduced, even after extirpation of the central nervous system, or in pieces of gut isolated in Ringer. Such movements are therefore independent of the central nervous system, of the nerve plexus in the body wall, and of the pharyngeal plexus, and they recall the autonomous movements of the vertebrate gut described by Cannon (1906), Pavlov (1910), Rubaschoff (1912), etc. Similar autonomous gut movements have been described in other annelids (*Arenicola*, *Nereis* and *Glycera*) by Wells (1937 a, b).

There are two possible explanations of these persistent co-ordinated rhythmic movements in the gut of earthworm; either they are the result of local reflexes through the nerve plexus in the gut wall, or they have a purely myogenic origin. If they are dependent on a local reflex, selective paralysis of the nerve elements by the use of a suitable poison or anaesthetic should abolish the movement. Experiments show that they do so. If the oesophagus, exposed by opening up the body wall, is painted with 5% novocaine, care being taken to restrict the drugs to the gut wall, loss of tone and abolition of peristalsis follow (figure 1). This suggests a dependence on a local reflex, but it does not demonstrate this conclusively, because it is impossible to be certain that the drugs have acted only on nervous tissue, and not also on the muscles.

Closer examination of the oesophagus enables the correspondence between its movements and those of the vertebrate gut to be extended further. Bayliss & Starling (1899) showed that stimulation of any portion of the mammalian small intestine causes contraction above this point and relaxation below it. They believed this reaction to be a purely local reflex carried out through the nerve plexus of the gut wall. The reflex was subsequently termed the 'myenteric reflex' by Cannon. An excised portion of the earthworm's oesophagus also exhibits the myenteric reflex when stimulated by pricking or squeezing with forceps.

INDIRECT CONTROL EXERCISED BY SEPTAL NERVES

If the septa running from the body wall to a portion of the gut such as the oesophagus are cut across, the latter undergoes a marked decrease in diameter accompanied by greatly decreased movement. This decrease in size may be due to one of two factors. First, the gut may simply 'collapse' when the supports attaching it to the body wall are removed. This possibility is removed by replacing the septa by mechanical supports.

Worms were opened up so as to expose the oesophagus, and a series of fine wire struts, running from the oesophagus to the body wall alongside each septum, were fixed in position under a binocular microscope. With the aid of a micrometer eyepiece, it was possible to fix the struts so that they stretched the oesophagus very slightly. When the septa were cut, despite the mechanical support provided by the struts, the oesophagus showed a marked decrease in diameter and appeared moniliform owing to the shrinkage of those portions in between successive struts.

This indicates that the shrinkage of the oesophagus with section of the septa is due to an active increase in the tone of its muscular wall. We must therefore consider the second possibility, namely, that the septa may carry nerves to the oesophagus, and that the observed change in tone is due to interference with this nerve supply. Such an effect would be quite similar to that following section of the vagal nerves to the mammalian stomach (Cannon 1906, p. 432).

The increase in tone might be produced in two ways, either by injury discharges in augmentor septal nerves produced by cutting, or by interruption of tonic impulses in septal inhibitor nerves. A simple experiment will decide between the two alternatives.

Several worms were opened up by a median dorsal incision in the anterior intestinal region; the septa were cut so that in this region the gut lay quite free in the coelom. As the worms were fresh from the soil, the gut contained much earth. The incision was closed, stitched, and the worms were replaced in moist soil. During the following 2 days the worms moved about normally and appeared quite healthy, but on the third day they began to show signs of weakness. They were then opened up again. It was found that the portion of the intestine separated from the body wall was still tightly constricted, and the passage of earth through the gut had been impeded, causing the portion immediately anterior to

the denervated region to become distended with accumulated earth passed down from more anterior regions, while the region behind the denervated zone was completely evacuated.

Although the injury discharge in nerves may persist for several hours, as in the sensory fibres of certain mammals (Adrian 1930), it is unlikely that they would persist for 3 days, and we are therefore led to believe that the increase in tone is due to interruption of tonic inhibitor impulses in septal nerves.

Evidence therefore indicates that the gut is maintained in a normal condition by tonic inhibitor impulses in septal nerves. This agrees with the results obtained by Wu (1939*a*), who, by means of experiments involving electrical stimulation of the central nervous system, comes to the conclusion that the septa carry inhibitor fibres. Further evidence in support of this conclusion will be given later.

BLOCKING OF TONIC INHIBITOR IMPULSES BY ANAESTHETICS

If 0.02–0.05 c.c. of a 10% solution of novocaine in Ringer is injected into the body-wall muscles of earthworm, a rise in tone of the neighbouring regions of the gut ensues (figure 2). The same quantity of 0.4% cocaine has a similar effect. Since novocaine exerts a direct effect on the gut (p. 274), it is imperative that in these experiments no fluid should escape from the body wall. Accordingly, the area receiving the injection was watched carefully under a microscope, and if any fluid was seen to escape from the body wall the preparation was rejected. Thus the effect on the gut can only be indirect, and since most anaesthetics commonly used on vertebrates are found to exert their characteristic effects on a large number of invertebrates (Harris 1900; Oestergren 1903; Kreckler 1910; Gray 1935, etc.), the above effect is most simply explained as due to blocking of tonic inhibitor impulses coming from nerve elements situated in the body wall.

It may be argued, however, that the rise in tone was due to mechanical stimulation of augmentor nerves through the act of injection. This is not so, because the injection of Ringer's fluid into the body wall under exactly the same conditions is without significant effect.

BLOCKING OF TONIC AUGMENTOR IMPULSES BY ANAESTHETICS

If instead of injecting novocaine into the muscles of the body wall it is painted on the peritoneum lining the body wall, care being taken to avoid direct contact of the drug with the gut, a marked fall in tone of the neighbouring regions of the gut ensues (figure 3). Such an effect indicates the existence of another set of nerve elements in the peritoneum of the body wall, which exert a tonic augmentor effect on the gut.

Previous experiments have provided evidence of nerves in the body wall and septum, which exert a tonic inhibitor effect on the gut, and, in addition, there is

now an indication of augmentor nerves in the body wall, but as yet no evidence of their presence in the septa. Such evidence is forthcoming from a study of the effect of a cold block on the septa.

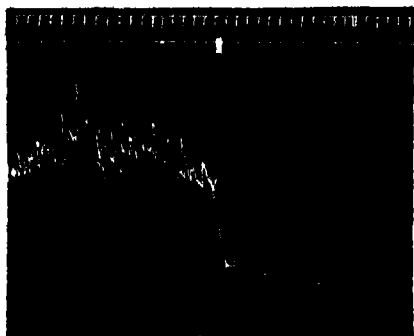


FIGURE 1

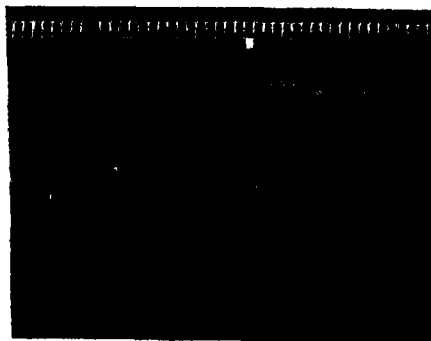


FIGURE 2

FIGURE 1. Abolition of spontaneous rhythmical contractions of gut wall. 5% solution of novocaine painted (at signal) on to the oesophagus (see p. 274). *Note.* In this, as in all succeeding figures, the records should be read from left to right, and the time scale is in half-minutes.

FIGURE 2. Blockage of inhibition by septal nerves. Effect on crop of 10% novocaine injected (at signal) into the muscles of neighbouring body wall (see p. 276). Drum stopped 3 min. during injection.

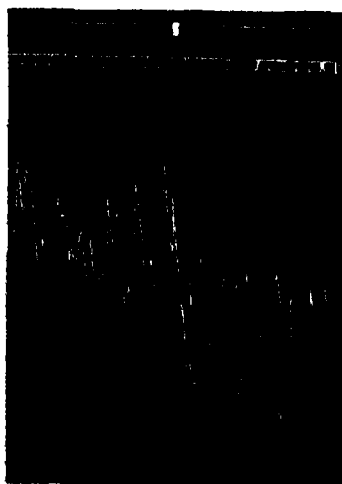


FIGURE 3. Blockage of augmentor impulses in peritoneal nerves. 10% novocaine painted (at signal) on the peritoneum lining the body wall surrounding the crop.

THE EFFECT OF COLD BLOCK

The method of applying a cold block is shown in figure 4. A double loop of fine capillary tubing (*c.t.*) was made to fit conveniently in the exposed coelom. Using an adjustable clamp, the capillary loops were fixed in contact with a septum (*sep.*)

under a binocular microscope, care being taken to avoid direct contact with the gut (*g.*). By using worms opened up in the mid-dorsal and mid-ventral lines, it was possible to rest the capillary tubing against the dorso-lateral and ventro-lateral regions of a septum. The capillary tubing was connected to a large syringe containing saline, immersed in a freezing mixture; and by depressing the plunger, fluid at about -2°C was circulated through the tubing. The movements of the portion of the gut nearest the septum were observed and recorded.

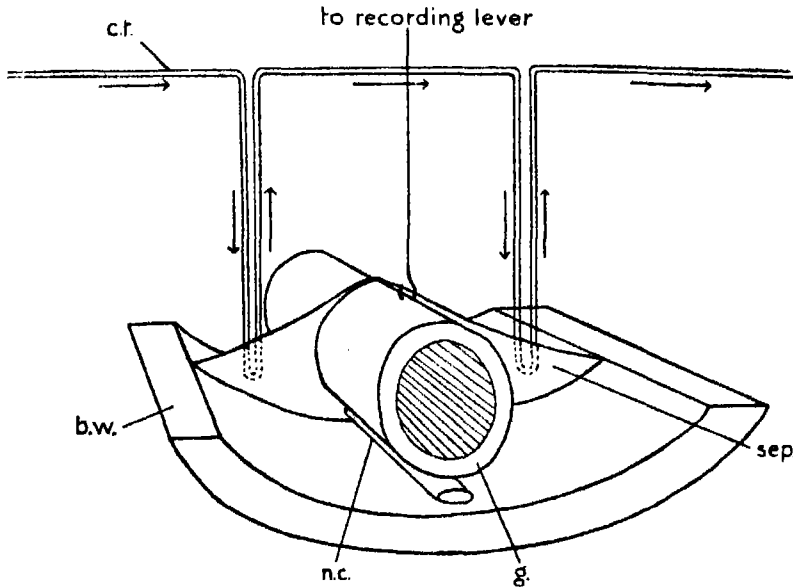


FIGURE 4. Diagram showing method of applying cold block (see p. 277). *b.w.* body wall of worm; *c.t.* capillary tubing through which saline is circulated as shown by arrows; *g.* gut; *n.c.* nerve cord; *sep.* septum.

Using this method, it becomes evident that the results of the experiment depend entirely on the region of the septum to which the cold block is applied. Several experiments were made, and all gave consistent results. The effects are clear to direct observation, but in view of the difficult technique it was not found possible to produce striking tracings.

If the cold block is applied to the dorso-lateral region of the septum, an increase in tone of the gut is observed. Since the application of sufficient cold to a nerve prevents the conduction of impulses in it, such experiments confirm the conjecture (p. 275) that the septa carry nerves exerting a tonic inhibitor effect on the gut, and show that these nerves are localized in the dorso-lateral regions of the septum. If, however, the cold block is applied to the ventro-lateral region of a septum, a fall in tone of the gut ensues. This suggests that the septa also carry augmentor nerves, which traverse their ventro-lateral aspect. This is significant because evidence of the existence of augmentor nerves in the body wall has already

been given, and there is the possibility (as in the case of the inhibitor nerves) that the two groups of nerves belong to the same system.

If we accept the argument that the effect of cutting the septa, anaesthetizing the body wall (both muscles and peritoneum), and applying a cold block to the septa, are all due to interruption of nerve impulses passing from the body wall and septa to the gut, then we reach the following conclusions. First, the alimentary canal is indirectly controlled by nerves from an outside source. Secondly, these nerves fall into two groups (augmentor and inhibitor), each localized in a special area of the body wall and septum. These conclusions can be tested directly by stimulating the augmentor and inhibitor nerves electrically.

ELECTRICAL STIMULATION OF THE AUGMENTOR AND INHIBITOR NERVES

The stimulating circuit used in the following experiments was of two types. In one, an a.c. current of 4 V drawn from a mains transformer and controlled by a 400 Ω potentiometer was used. In the other, an induction coil in circuit with a 2 V battery was employed. The electrodes used were of fine silver wire. Insulation was found to be unnecessary, since it was always possible to manipulate the electrodes so that only the tips were in contact with the tissues. In order to prevent short-circuiting and spread of current the preparations were kept carefully drained of superfluous Ringer.

Stimulation of any portion of the peritoneal surface of the body wall is followed by a marked increase in tone of the neighbouring portions of the gut. This effect is shown for the oesophagus in figure 5. The increase in tone is not confined to that region of the gut opposite the portion of the peritoneum which is stimulated, but may extend over a large number of segments (see below), and though always well marked, its extent and duration vary greatly in different individuals. The experiments in which novocaine was painted on the peritoneum also indicate the existence of augmentor nerves in the peritoneum of the body wall.

However, before the results of the above experiments can be fully accepted, certain objections must be met.

First, since the muscles of the body wall also contract when the peritoneum is stimulated, it is clear that these might contribute to a greater or lesser extent toward the rise in tone registered by the recording apparatus. This possibility was eliminated by securing the body wall very firmly with pins, after which it could be shown that contraction of the body wall exerted no significant effect on the recording lever.

Secondly, since the septa are muscular they could influence the recording lever by pulling on the gut, as already suggested by Wu (1939*a*). This possibility is enhanced by the radial arrangement of most of the septal muscles described by de Ribaucourt (1901), which means that in the type of preparation used in these experiments (p. 274) the septal muscles could pull the gut downwards (i.e. farther

into the coelom), and by suddenly changing the base line of the records in this way they could simulate a rise in tone of the gut. However, this possibility could be eliminated by observation, for if the effect shown in figure 5 is due to contraction of radial septal muscles, the gut would be subject to a lateral as well as a downward pull, and would appear dilated at the points where the septa meet the gut, which is not the case.

The convenient size and rapid response of the oesophagus make it very suitable for observation, and the question of possible participation of septal muscles in the effect was studied in detail in this region. In all experiments, stimulation of the peritoneum (or certain augmentor nerves to be described later) resulted, not in an increase, but in a decrease in diameter of the oesophagus, and in several such experiments its diameter (measured by a micrometer eyepiece) was seen to shrink to about half following stimulation. It is significant that the shrinkage occurs at the junctions of the septa and oesophagus as well as in the intervening regions. It is therefore obvious that the radial septal muscles cannot be directly responsible for the effect.

However, these muscles might have an indirect effect. Owing to the disposition of the septal muscles in the oesophageal region, it is possible for the gut to be pulled forward by their contraction, but being anchored by the septa of more posterior regions, the gut would be stretched, and because it is elastic, appear to become narrower. This does not explain the behaviour of the oesophagus however, because in many experiments narrowing occurred without any appreciable forward movement, and even when this was observed, it was too small to account for the simultaneous shrinking of the oesophagus.

The only type of septal muscle that could produce narrowing of the oesophagus is a sphincter muscle. No such muscles have been described, and I have searched for them in serial sections (transverse, longitudinal and frontal), but I found none. Septal muscles which are disposed in a circular fashion are described by de Ribaucourt, but they are feebly developed and cannot be part of a sphincter round the gut, as they are stated to be absent from the ventral portions of the septa. Furthermore, being situated at a considerable distance from the gut, they can easily be put out of action by cutting across the dorsal regions of the septa. Such an operation does not abolish the response of the oesophagus already described. We may therefore conclude that the narrowing of the gut which follows stimulation of the peritoneum and augmentor nerves (see below) represents an increase in tone brought about by the gut muscles, and not by the muscles in the septa.

The conclusion is confirmed by the experiment shown in figure 6, in which 1% novocaine had been painted on the septa (at arrow), with the result that the rise in tone of the oesophagus produced by stimulating the septum was abolished (signals 3 and 4). Now novocaine is known to prevent the response of an effector to stimulation by blocking the passage of impulses in its nerve supply, and not by acting directly on the contractile mechanism of its muscles. Hence assuming that the drug acts in its characteristic manner, it is unlikely that it would prevent

the contraction of the septal muscles in response to direct stimulation. Therefore it is unlikely that the septal muscles are responsible for this effect.

The third objection to the type of experiment shown in figure 5 is that the extremely small space between the body wall and gut and between the electrodes placed on the body wall and the nearest septa, makes it possible for the stimulating

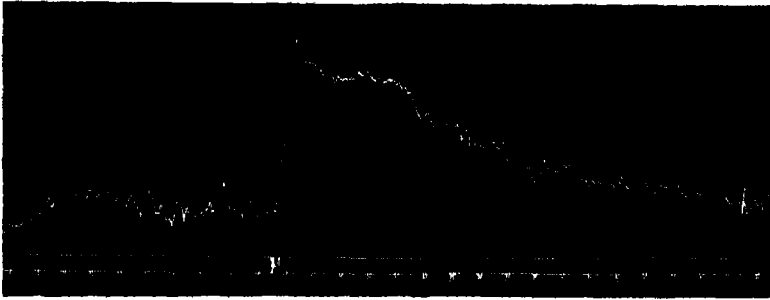


FIGURE 5. Stimulation of augmentor nerves in peritoneum. Effect on tone of oesophagus of electrical stimulation (at signal) of the peritoneum lining the neighbouring body wall (see p. 279).

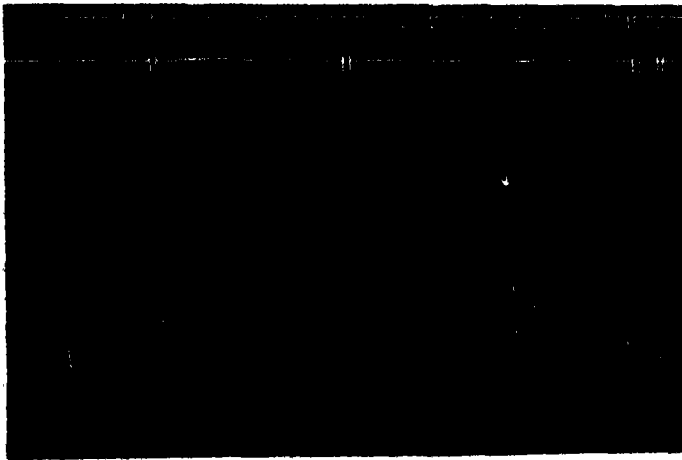


FIGURE 6. Control in which participation of septal nerves is prevented by anaesthetic (see p. 280). Signals 1 and 2, stimulation of septa. Signals 3 and 4, repetition of stimulation after painting septa with 1% novocaine.

current to spread to the oesophagus without participation of any nerves, resulting in direct excitation of the oesophageal muscles. This possibility may be tested in the first place indirectly, by making use of the remarkable conduction of augmentor impulses to distant portions of the gut. If the peritoneal surface of the body wall is stimulated as far as thirteen segments away from the portion of the gut which is attached to the recording lever, a marked increase in tone results (figure 7). This is still produced after transection of the gut (figure 8). Such experiments

eliminate the possibility of the effect being produced by nerve impulses or current spreading through the gut wall, or the purely mechanical pull exerted on a passive region by increase in tone of the gut nearest the electrodes.

Another possibility is that the central nervous system lying close to the body wall may be excited by spread of current in the stimulated region, and may transmit the excitation to distant regions of the gut which are attached to the recording lever. This is eliminated by removing the nerve cord in the region of stimulation and repeating the experiment shown in figure 7, when it will be seen that the rise in tone is still produced (figure 9), showing that the central nervous system is not involved.

Thus excitation must be conducted from the point of stimulation by nerve elements situated in the peritoneum of the body wall, and reaches the oesophagus via nerves in the septa.

A direct test for possible spread of current by short-circuiting may be made by observing the effect of stimulating the peritoneal surface of the body wall nearest to the region of the gut which is attached to the recording lever, after the neighbouring septa have been cut across, or alternatively treated with Ringer containing novocaine. If the augmentor response seen in figure 5 involves septal nerves, then cutting across the septa or treating them with an anaesthetic should abolish it. If, on the other hand, the response is merely the result of direct stimulation owing to short circuits set up in the moist tissues, then cutting or treating the septa with anaesthetic should have no such effect; in the latter case, on the contrary, the response should be elicited more readily, because in these experiments the septa were continually moistened with Ringer containing novocaine.

In worms which have been opened up mid-dorsally and in which the septa of the oesophageal region have been cut across, leaving only a small mid-ventral portion of each septum intact, stimulation of the peritoneum of the neighbouring body wall has no effect on the tone of the oesophagus. However, such experiments alone cannot be accepted as excluding the possibility that the rise in tone may be due to spread of current. It must be borne in mind that, since the recording lever exerts a slight pull on the region of the gut to which it is attached, particularly as the support normally provided by the septa has been largely removed by operation, it is possible that the gut may be pulled away from the body wall in the region of stimulation, and hence a slight spread of current under these conditions would be without effect. This objection may be overcome by holding the gut in contact with the body wall by means of light struts and then repeating the experiment. As figure 10 shows, stimulation of the peritoneum at all signals still evokes no significant response in the gut, even when the electrodes are placed very close to the gut without actually touching it. Thus short-circuiting of current cannot be held responsible for the rise in tone of the gut following stimulation of the peritoneum, which therefore must be due to impulses in augmentor nerves in the septa. This conclusion is fully confirmed by experiments in which the septa and peritoneum (see below) are treated with anaesthetics. As figure 6 shows,

painting the septa with 1 % novocaine prevents the rise in tone of the oesophagus which normally follows stimulation of the peritoneum.

The above experiments, in addition to excluding the possibility of a short-circuit being responsible for the augmentor response, also show that it involves

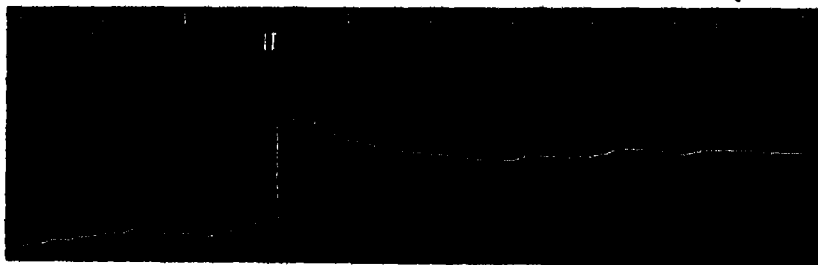


FIGURE 7. Spread of augmentor impulses through the peritoneum. Effect on tone of oesophagus of electrical stimulation (see p. 279) at signal, of the peritoneum lining a distant region of the body wall (p. 281).

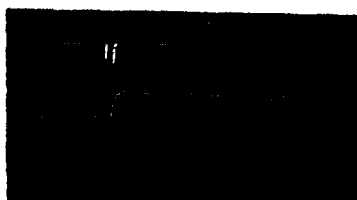


FIGURE 8

FIGURE 8. Experiment showing that the rise in tone in figure 7 is not due to conduction of impulses through the gut wall (p. 281).



FIGURE 9

FIGURE 9. Experiment showing that the response in figure 7 does not involve the central nervous system (p. 282).



FIGURE 10. Control showing that the rise in tone produced by electrical stimulation (p. 279) is not due to short-circuiting (p. 282); the electrodes were placed on the peritoneum lining the body wall as near as possible to the gut, after the septa had been cut across.

nerve elements, and that they are situated in the septa. Using a similar technique, it can be shown that the response also involves nerve elements situated in the peritoneum. Thus the rise in tone of the crop which normally follows stimulation of the peritoneum lining the neighbouring body wall can be abolished by treating the peritoneum with 1 % novocaine. This action of novocaine can be explained

satisfactorily only by assuming that it has put out of action nerve elements located in the peritoneum. Direct confirmation of this can be obtained by removing the peritoneum from an area of the body wall which is subsequently stimulated as in previous experiments. The peritoneum must be dissected away from the underlying muscles with extreme care under a binocular microscope. As figure 11 shows, after such an operation, the rise in tone normally produced by electrical stimulation of the body wall (1st signal) is abolished by removing the peritoneum (2nd signal). Thus we now reach the definite conclusion that the septa and peritoneum lining the body wall contain nerve elements which by their action increase the tone of the alimentary canal.

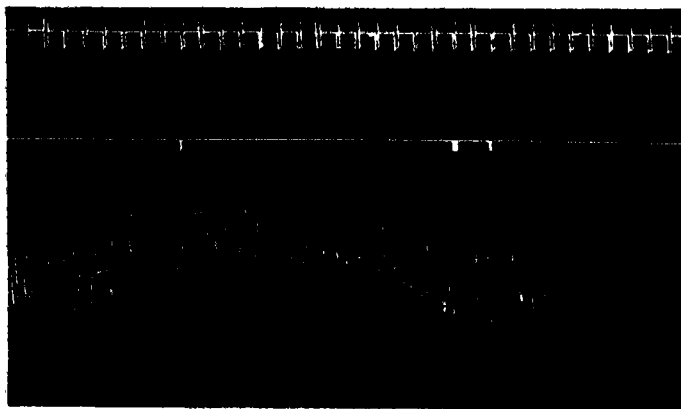


FIGURE 11. Experiment showing that the removal of the peritoneum lining the body wall abolishes the augmentor response (p. 284).

Experiments in which novocaine was injected into the body-wall muscles, and on the application of a cold block to the dorso-lateral regions of a septum, supplied evidence of a system of nerve elements situated in the body-wall muscles and dorso-lateral region of the septa, which exerted an inhibitor effect on the tone of the gut. The evidence may be confirmed by electrical stimulation as in the case of the augmentor nerves. If, instead of placing the electrodes on the peritoneum, they are sunk in the muscles at the cut edge of the body wall, stimulation is usually followed by a fall in tone of the neighbouring portions of the gut. Occasionally, a rise in tone is produced, but this can be changed to a fall by stripping the peritoneum from the stimulated area, and appears therefore to be due to spread of current from the muscles to the overlying peritoneum, with consequent masking of the inhibition. This confirms the conjecture that the body wall contains a second set of nerve elements, situated more deeply in the musculature, and exercising an inhibitor effect on the gut.

The significant results of the experiments which involved electrical stimulation may now be summarized as follows. They demonstrate directly the presence of both augmentor and inhibitor nerves in the body wall, and have shown the locali-

zation of the augmentor nerve elements in the peritoneum and inhibitor nerves in the muscle layer of the body wall. The augmentor nerves in the peritoneum and at least some of those in the septa have been shown to belong to the same system, and the spread of impulses through the peritoneum indicates that its augmentor nerve elements are probably arranged in the form of a plexus.

Now that the relative position of the augmentor and inhibitor nerve elements in the body wall has been found, it remains to be shown how the two are situated relative to one another in the septa. This is revealed when the relation of both systems to the central nervous system is studied.

RELATION OF THE PLEXUS SYSTEM TO THE CENTRAL NERVOUS SYSTEM

The body wall of the earthworm has long been known to receive its nerve supply from the central nervous system by way of the segmental nerves; if, therefore, there is any connexion between the nerve elements in the body wall which have been shown to influence the gut, and the central nervous system, it should be possible to bring about changes in tone of the gut by stimulating the segmental nerves.

The nerves were exposed for stimulation as follows. Worms were opened up slightly to one side of the mid-dorsal line. On pinning out the cut edges of the body wall, the nerve cord and segmental nerves were carried over to a lateral position where, by carefully stretching out the body wall, the nerves were conveniently exposed so that the electrodes could be manipulated with a minimum risk of contact with surrounding structures. In many cases it was found convenient to mount the preparation under a low-power binocular and, picking up the nerve on one electrode held by a micro-manipulator, to stimulate it by touching the remaining free portion with the other electrode manipulated by hand.

If the single (anterior) segmental nerve of either side be stimulated, the corresponding region of the gut shows a distinct fall in tone (figure 12). Great care is necessary, however, for should the electrodes touch any portion of the surrounding peritoneum, the effect is completely masked by a powerful rise in tone. Since the fall in tone can be prevented by cutting across the nerve at the point where it enters the body wall, or by cutting across the neighbouring septa, it is clear that the *anterior* segmental nerves contain fibres conveying inhibitor impulses which pass by way of nerve elements in the body wall and septa to the neighbouring portions of the alimentary canal. These results harmonize with those obtained by Wu (1939a), who stimulated the anterior portion of the central nervous system underlying the crop and gizzard and described a fall in tone of these portions of the gut. From the description given by Wu (p. 194) it is evident that, although he isolated that portion of the central nervous system beneath the crop and gizzard, the segmental nerves of the region remained intact, and thus the fall in tone could be produced by inhibitor fibres in these nerves.

In contrast to the above, stimulation of the median and posterior segmental

nerves produces a rise in tone of the portion of the gut situated in the same segment. The same effect is produced by stimulating either the median or posterior nerve alone, but in the case of the gizzard the rise in tone is produced only by the posterior nerve, stimulation of the median nerve having no effect. In the latter experiments, since the two nerves lie very close together, in order to eliminate the possibility of the effects being due to accidental stimulation of both nerves, one of the pair was cut across at its point of entry into the body wall.



FIGURE 12



FIGURE 13

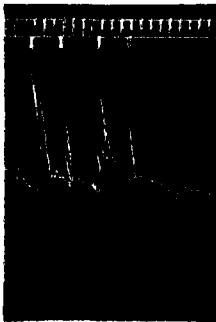


FIGURE 14



FIGURE 15

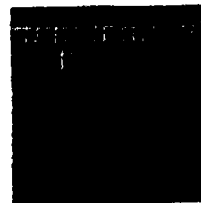


FIGURE 16

FIGURES 12-16. Responses of crop and intestine to electrical stimulation of segmental nerves (stimulating circuit as described on p. 279). Figure 12 (crop). Effect of stimulating anterior segmental nerve (at signal). Figure 13 (crop). Effect of stimulating median and posterior segmental nerves (at signal). Figure 14 (crop). Effect of stimulating isolated median and posterior segmental nerves (all signals). Figure 15 (crop). Effect of stimulating median and posterior segmental nerves (at signal) after ventral portion of septum had been cut across (p. 288). Figure 16 (intestine). Similar response of intestine under same conditions as figure 15.

Exactly as in the case of the anterior segmental nerves, the effect produced on the gut by stimulation of the median and posterior nerves is abolished by cutting across the neighbouring septa, or by transecting the nerves before their entry into the body wall. It follows therefore that the median and posterior nerves (posterior only, in the case of the gizzard region) convey impulses which, passing by way of

nerve elements situated in the body wall and septa, augment the tone of neighbouring portions of the gut.

It may be argued, however, that in coming to the above conclusion we have made the unjustified assumption that the nerve elements involved in the response are purely efferent and that the impulses may have passed to the central nervous system by way of afferent fibres in the segmental nerves and thence to the gut by some quite different route. The following experiment illustrates the truly efferent character of the nerves. The median and posterior nerves on one side of one segment were completely isolated so that the impulses could reach the gut by one route only, viz. through the nerves to the body wall and to the gut via the septa. A rise in tone of the neighbouring portion of the gut followed stimulation of the paired nerves (figure 14), just as if the central nervous system had been intact, and thus the response is dependent on impulses in efferent nerves passing along the route described. Thus it follows that augmentor, as well as inhibitor, nerves pass from the central nervous system to the gut, via the body wall and septa.

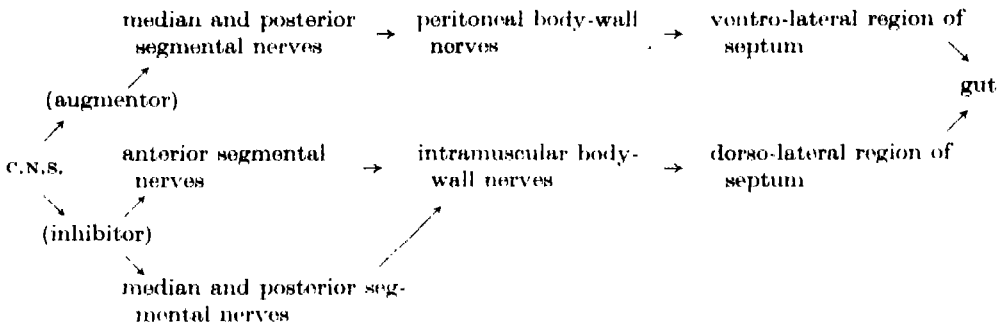
Wu (1939*a*) describes an increase in tone of the crop produced by stimulating the central nervous system, but believes it is apparent rather than real, and due to the contraction of the septa. He does not mention any evidence of augmentor fibres in the segmental nerves or septa.

Since the distribution of augmentor and inhibitor nerve elements among the segmental nerves is now known, it becomes a simple matter to discover by what route augmentor and inhibitor impulses traverse the septa after leaving the central nervous system. If after a portion of the septum has been removed, or cut across, stimulation of one of the neighbouring segmental nerves fails to elicit the normal response, then clearly the portion of the septum operated on carries the necessary nerve elements.

Separation of the mid-dorsal region of the septum from the body wall does not greatly affect the response to stimulation of the segmental nerves. The same applies to the mid-ventral region. Hence the mid-dorsal and mid-ventral regions of the septum cannot be the main pathways of the augmentor and inhibitor nerves. If, however, the dorso-lateral region of the septum is separated from the body wall, stimulation of the anterior segmental nerve no longer has any effect. Thus the inhibitor nerves run through the dorso-lateral region of the septa. When the ventro-lateral region of a septum is separated from the body wall, stimulation of the median and posterior nerves of the segment brings about a fall in tone of the adjacent region of the gut (figures 15, 16), instead of the usual rise. This is true for all regions of the gut except the gizzard, where such treatment merely abolishes the rise normally produced by stimulating the median and posterior nerves. Thus the augmentor nerves to all regions of the gut run through the ventro-lateral regions of the septa and, in addition, the latter experiments show that the median and posterior segmental nerves (with the exception of those in the gizzard region) carry inhibitor as well as augmentor nerve elements.

It thus follows that if the fall in tone of the crop following stimulation of the central nervous system described by Wu (1939*a*) is produced by impulses passing through segmental nerves and septa, it would seem to be either a complex effect in which the action of the inhibitors predominates, or the arrangement of the inhibitor elements in the central nervous system must be such that, under the conditions of his experiments, only the inhibitors were stimulated.

The results of the foregoing experiments may be summarized thus:



The distribution of septal nerves described above agrees completely with experiments in which a cold block was applied to the septa (p. 278), and confirms the conjectures made from these experiments. These results are summarized in figure 17.

INDIRECT CONTROL EXERCISED BY THE PHARYNGEAL PLEXUSES

Situated on either side of the anterior aspect of the pharynx is a plexus of nerve elements lying partially embedded in the pharyngeal musculature. It is connected with the central nervous system by several short rami which pass to it from the peripharyngeal commissures.

The plexus has been described by various workers as the 'pharyngeal plexus', 'pharyngeal ganglion', 'ganglionated thickening of the enteric nerve' and 'enteric nerve'. It is proposed to use the term 'pharyngeal plexus' in this account, as it is descriptive of its form and, at the same time, avoids confusion with the general nerve plexus in the gut wall, of which it may be considered to be a special portion (Dechant 1906). The essentials of its structure have been described by several workers including Quatrefages (1847), Clark (1856), Dechant (1906), and Hess (1925).

The relation of the plexus to the central nervous system in *Pheretima communissima* has been described by Ogawa (1934, 1939), who mentions efferent nerve fibres running through the pharyngeal plexus to the alimentary canal, having their cells situated in the cerebral and suboesophageal ganglia. He also describes afferent nerves passing from the gut wall, and terminating in the pharyngeal plexus.

In view of the importance of this structure, its anatomy was investigated and reconstructed from serial sections, and its significance in the indirect control of the gut musculature was examined by the experiments to be described.

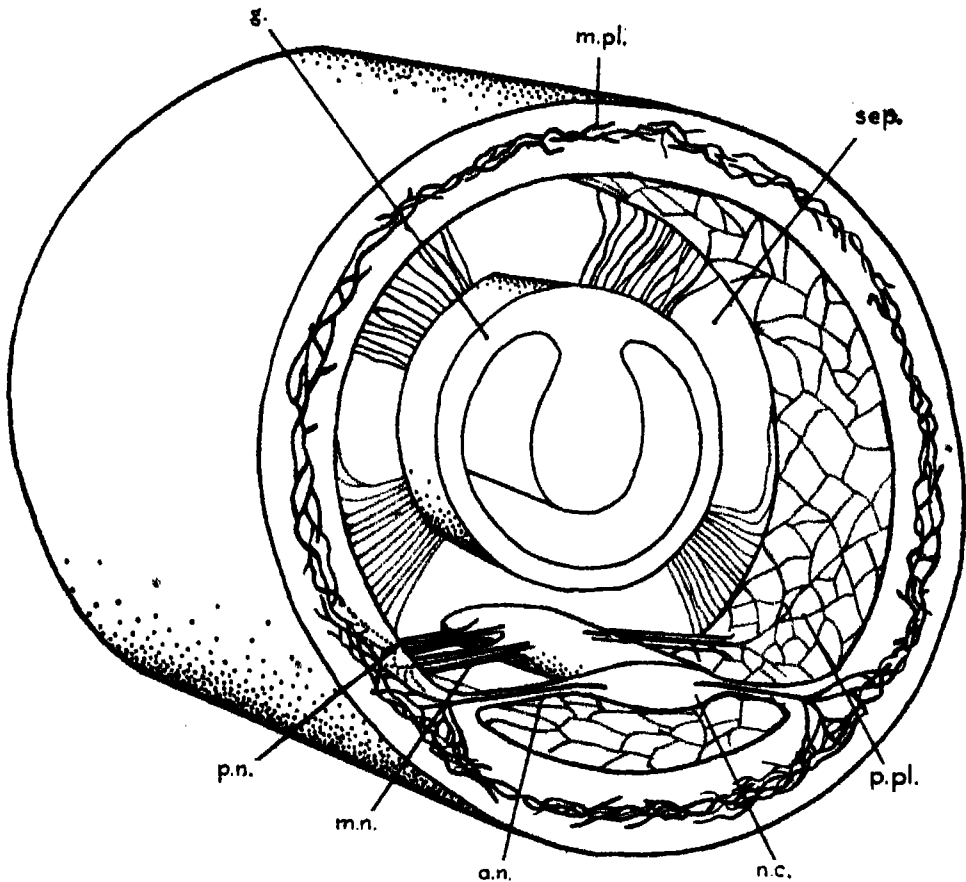


FIGURE 17. Diagram of a thick slice of earthworm showing distribution and relationships of nerves supplying the alimentary canal. Augmentor nerves red, inhibitor nerves black. N.B. The illustration of peritoneal and intramuscular nerves is purely conventional, no attempt is made to indicate the precise structure and relationship of the nerve elements. *a.n.* anterior segmental nerve; *g.* gut; *m.n.* median segmental nerve; *m.pl.* intramuscular nerves; *n.c.* ventral nerve cord of central nervous system; *p.n.* nerve elements leaving posterior segmental nerve to join the intramuscular nerve elements; *p.pl.* nerves in peritoneum lining the body wall; *sep.* septum.

STRUCTURE OF THE PHARYNGEAL PLEXUS (see figure 18)

Methods. The anatomy was reconstructed from serial sections by means of a micro-projection apparatus. Material was fixed in Petrunkevitch, Duboscq-Brasil, and Carnoy's fluids. Sections were cut at 10μ in the transverse, sagittal-longitudinal, and frontal planes.

The plexus on each side consists of a large ganglionic mass (*m.*) arising from the peripharyngeal commissure (*p.c.*) of the same side by seven roots (*r. i.*, *r. ii.*, etc.). From the main mass numerous large nerves (*n.*, *n.¹*) pass into the wall of the pharynx and join a plexus ramifying in the epithelium and subjacent musculature of the pharynx which make connexion with similarly disposed nerve elements extending posteriorly along the gut. A more detailed description of the disposition and structure of the plexus in the gut wall is reserved for a future account.

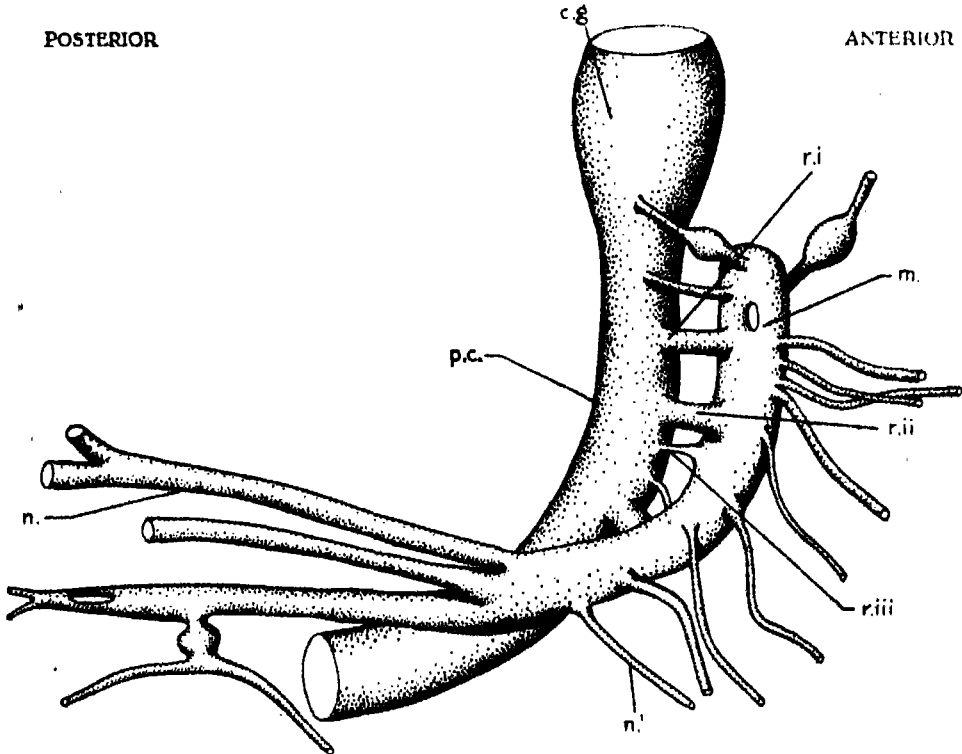


FIGURE 18. Reconstruction of pharyngeal plexus of left side, showing its relations with the left peripharyngeal commissure of the central nervous system. The plexus is viewed as seen from inside the pharynx. *c.g.* left cerebral ganglion; *m.* main ganglionic mass of the plexus; *n.*, *n.¹*, nerves passing into wall of pharynx; *p.c.* left peripharyngeal commissure; *r. i.*, *r. ii.*, *r. iii.*, nerve roots by which plexus arises from peripharyngeal commissure.

THE CONTROL EXERCISED BY THE PHARYNGEAL PLEXUS

Methods. The nature of the control over the tone of the alimentary canal exercised by the pharyngeal plexus was determined by stimulating the plexus electrically. The stimulating circuits used were the same as those described on p. 279.

Owing to the small size of the plexus, and its intimate association with the wall of the pharynx, it proved exceedingly difficult to stimulate the plexus without also stimulating the wall of the pharynx directly.

The plexus of one side was therefore exposed by dissection, and two exceedingly fine silver wires were passed between the main ganglionic mass (*m.* figure 18) and the wall of the pharynx. By gently raising and tautening the wires, the plexus and associated peripharyngeal commissure, which had been isolated (see below), was raised from the surface of the gut, and after ensuring adequate insulation of the wires, a stimulating current was sent through them, and its effect on the various regions of the gut determined. The plexus so elevated was kept moistened with Ringer, the excess fluid being withdrawn by a fine pipette to lessen the chances of a short-circuit occurring.



FIGURE 20

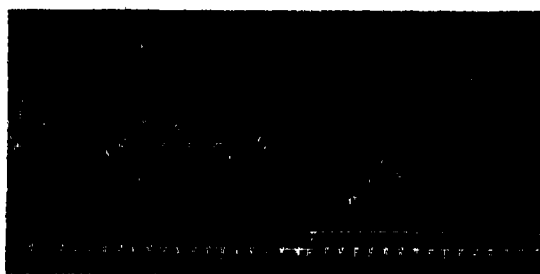


FIGURE 19



FIGURE 21



FIGURE 22

FIGURES 19-22. Responses of various regions of the gut to electrical stimulation (at signal) of the pharyngeal plexus. The rise in tone at 2nd signal in figures 20 and 21 is incidental, and due to stimulation of the peritoneum lining the body wall. Figure 19. Oesophagus. Figure 20. Crop. Figure 21. Gizzard. Figure 22. Intestine.

To ensure that nervous impulses did not reach the gut via the central nervous system, segmental nerves, and the body-wall plexuses, the peripharyngeal commissure of the same side was isolated from the rest of the central nervous system, by cutting it across at either end.

Electrical stimulation of the pharyngeal plexus produced a fall in tone of the oesophagus, crop, gizzard and intestine. These effects are shown in figures 19-22.

It should be noted, however, that unless great care was taken to avoid a spread of current to the wall of the pharynx, the effect on the whole gut was variable, very often a distinct rise in tone being produced.

Since the pharyngeal plexus contains nerve elements which join the nerve plexus in the gut wall (see above, and Dechant (1906)), it follows from the foregoing experiments that these elements conduct inhibitor impulses which must pass through the nerve plexus in the gut wall to regions as remote as the intestine. Wu (1939a), stimulating the circumpharyngeal commissures of earthworm, found that the crop underwent an increase in tone, and no mention is made of any inhibitor nerves entering the gut from the pharyngeal plexus. As stated above, in experiments in which there was a possibility of an accidental short-circuit being set up, the gut often responded to stimulation with a rise in tone. This may possibly indicate the presence of another group of nerve elements in the gut wall which conduct augmentor impulses along the gut, and so far as they indicate that possibility, the results are in agreement with those of Wu; but no evidence definitely indicating that such nerve elements are connected with the pharyngeal plexus was obtained from the experiments of this investigation. It should be noted, however, from the description given by Wu of his technique (p. 194), it is impossible to be certain that in his experiments impulses leaving the circumpharyngeal commissures could not reach the gut via the nerves arising from the cerebral and suboesophageal ganglia and nerve elements in the body wall and septa.

The fact that the pharyngeal plexus can exert a significant effect on the tone of the alimentary canal is in complete harmony with the description of the structure of the plexus already given, with the nervous connexion existing between the pharyngeal plexus and the alimentary canal shown by the work of Dechant (1906), and in particular by the work of Ogawa noted above.

Thus the alimentary canal of *Lumbricus* is indirectly controlled by two main groups of nerves, first, those which pass to the gut via the segmental nerves, body wall and septa, and secondly, those which pass via the pharyngeal plexus.

DISCUSSION AND CONCLUSION

It has been shown that the nerve supply to the gut of *Lumbricus* and *Allolobophora* arises from the central nervous system by two main channels. One supply is derived from the peripharyngeal commissures by way of the pharyngeal plexuses, the other from the whole length of the ventral nerve cord, via the segmental nerves. The supply via the pharyngeal plexuses has long been known and is described by all the earlier workers on earthworm. Clark (1856) and Lankester (1865) seem to regard it as constituting the entire nerve supply to the gut.

A somewhat similar nerve supply to the gut arising from the peripharyngeal commissures (or their equivalents) is described in other types of annelids by numerous workers, including Bristol (1898) in *Nepheleis*, Gravier (1898) in *Glycera*,

Ashworth (1909) in *Arenicola*, and Gaskell (1914) in *Hirudo*. Therefore the observations made in this account on the structure and relationships of earthworm fall into line with those of other workers. Nerve elements situated in the wall of the anterior portion of the gut of other annelids such as *Arenicola* have been shown by Wells (1937a) to play a specialized and important role connected with the burrowing habits of the animal. A somewhat similar mechanism is also described in *Glycera* (Wells 1937b).

The nerve supply from the ventral nerve cord is not mentioned by any of the earlier workers on *Lumbricus*, and furthermore, Bristol (1898) and Gaskell (1914), working on *Nepheleis* and *Hirudo* respectively, failed to find any nerves passing from the ventral nerve cord to the alimentary canal, despite careful search. The histological preparations described by Smallwood (1923), which are interpreted as showing the existence in earthworms of nerves passing to the alimentary canal from plexuses situated in the body wall, via the septa, are of greatest interest because the physiological evidence presented in the foregoing account agrees so closely with Smallwood's histological evidence. It seems therefore that unless the innervation of the gut from two main sources is peculiar to earthworms among annelids, the innervation of the alimentary canal of other annelids requires re-investigation, both from a histological and a physiological standpoint. It is significant in this connexion to note that the histological technique used by Bristol and Gaskell was not nearly so refined as that used by Smallwood, who used the Bielschowsky-Paton technique, and since the nerves described by Smallwood are very fine, the application of silver impregnation methods to other annelids might yield most valuable results. The writer proposes to attempt this at a later date.

The relation between the central nervous system, the nerves in the peritoneum and muscular layers of the body wall, and the septal nerves, together with the physiological significance of these peripheral nerve elements, is here described for the first time. The connexion between the central nervous system and the peripheral plexuses by means of fibres running in the segmental nerves is especially noteworthy because Gaskell has shown that the 'blood' (coelomic) vessels of *Hirudo* are also controlled by fibres from the segmental nerves. Such correspondence may possibly indicate a segmental organization of the autonomic nervous system of annelids, which would be of great interest in view of the evidence of segmental organization in the autonomic nervous system of vertebrates.

When the control exercised by the nervous system over the alimentary canal is compared in the earthworm and vertebrates, the similarities are remarkable. The alimentary canal in both cases exhibits autonomous rhythmic movements, which seem to be co-ordinated in earthworm as in the vertebrate (Bayliss & Starling 1899; Magnus 1904; Cannon 1912, etc.), by nerve elements in the gut wall. Furthermore, direct stimulation of the gut in earthworm results in a response strikingly similar to that of the 'myenteric reflex' (Cannon), exhibited by the small intestine of vertebrates. Again, the alimentary canal of earthworm, like that of vertebrates, is indirectly controlled by two groups of nerves from an

outside source, the augmentors and inhibitors, which are antagonistic in their effects, and at once recall the sympathetic and parasympathetic nerves of higher vertebrates. These striking similarities have their counterpart in *Hirudo* where the coelomic vessels are subject to indirect control by antagonistic systems of nerves (Gaskell 1914). The comparison extends further, for some evidence has been obtained that the nerves which decrease the tone of the gut of *Lumbricus* are adrenergic, while those which increase the tone of the gut appear to be cholinergic. This again recalls the condition found in higher vertebrates, in which, generally speaking, the adrenergic (post-ganglionic) sympathetic fibres bring about a decrease in tone of the stomach and intestines, while the cholinergic parasympathetic fibres bring about an increase in tone (see Dale 1933, 1934).

However, remarkable though these similarities may be, it must be borne in mind that they exist only as broad generalizations. Thus in *Lumbricus* the effects of the two groups of nerves on the tone of the musculature throughout the gut (apart perhaps from that of the pharynx, their actions on which have not been investigated) seems to be invariable. This is not the case in vertebrates as shown by investigations covering a wide range in vertebrate organization. Young (1933) describes the effect of parasympathetic (vagal) stimulation on the stomach of selachians as variable. Similar variable effects of vagal stimulation on the stomach are recorded by Dixon (1902) in frog, Carlson & Luckhardt (1920) in reptiles, Carlson, Boyd & Pearcey (1922 *a, b*), and McCrea, McSwiney & Stopford (1925), etc., in mammals. Dixon (in frog) and McCrea, McSwiney & Stopford (in mammals) find that the effect of parasympathetic stimulation on the stomach varies with the tone of the gastric musculature at the time of stimulation. Stimulation of the sympathetic (splanchnic) nerves has also a variable effect on the tone of the stomach (McCrea & McSwiney 1928).

Again, the effect on the tone of the intestine of selachians produced by stimulating the sympathetic nerve supply is described by Young (1933) as variable, and dependent on the tone of the intestinal musculature at the time of stimulation, while Bayliss & Starling (1899), and Carlson (1930), obtain similar effects on the tone of the intestine by stimulating both sympathetic and parasympathetic nerves in mammals.

So far as the innervation of the earthworm gut from the ventral nerve cord is concerned, the results point to a similar view to that expressed by several workers concerning the origin of the autonomic nervous system of vertebrates. Both seem to be differentiated out of a diffuse plexus of nerve elements such as constitutes the nervous system of the lower invertebrates. The participation of the body-wall plexuses in the efferent visceral nerve tracts of *Lumbricus* lends striking support to this view. Pending further investigation into the nerve supply of viscera other than the alimentary canal of earthworm, it is premature to attempt a more detailed discussion of this matter.

It is with great pleasure that I acknowledge the constant help and invaluable

advice given to me by Dr C. F. A. Pantin, F.R.S. My thanks are also due to Professor H. Graham Cannon, F.R.S., Professor J. Stanley Gardiner, F.R.S., Professor J. Gray, F.R.S. (who kindly granted me facilities for research in the Zoology Department, Cambridge), Dr H. Blaschko, and Dr D. Richter.

The research was undertaken whilst holding a Rouse Ball Research Studentship awarded by Trinity College, Cambridge.

REFERENCES

- Adolf, E. F. 1927 *J. Exp. Zool.* **47**, 31.
- Adrian, E. D. 1930 *Proc. Roy. Soc. B*, **106**, 598.
- Ashworth, J. H. 1909 *Phil. Trans. B*, **200**, 427.
- Bayliss, W. M. & Starling, E. H. 1899 *J. Physiol.* **24**, 99.
- Bristol, H. 1898 *J. Morphol.* **15**, 17.
- Cannon, W. B. 1906 *Amer. J. Physiol.* **17**, 429.
- Cannon, W. B. 1912 *Amer. J. Physiol.* **30**, 114.
- Carlson, A. J. 1930 *J. Amer. Med. Ass.* **94**, 78.
- Carlson, A. J., Boyd, T. E. & Pearcey, J. F. 1922a *Amer. J. Physiol.* **61**, 14.
- Carlson, A. J., Boyd, T. E. & Pearcey, J. F. 1922b *Arch. Int. Méd. Exp.* **33**, 281.
- Carlson, A. J. & Luckhardt, A. B. 1920 *Amer. J. Physiol.* **54**, 261.
- Clark, J. L. 1856 *Proc. Roy. Soc.* **8**, 343.
- Dale, H. H. 1933 *J. Physiol.* **80**, 10P.
- Dale, H. H. 1934 *Brit. Med. J.* **1**, 835.
- Dechant, E. 1906 *Arb. Zool. Inst. Univ. Wien*, **16**, 361.
- Dixon, W. E. 1902 *J. Physiol.* **28**, 57.
- Gaskell, J. F. 1914 *Phil. Trans. B*, **205**, 153.
- Gray, P. 1935 *Watson's Microscope Record*, No. 35. London.
- Harris, G. T. 1900 *J. Roy. Micr. Soc.* **20** (ser. 2), 404.
- Hess, W. N. 1925 *J. Morph.* **40**, 235.
- Krecker, F. H. 1910 *Z. wiss. Zool.* **45**, 383.
- Lankester, E. R. 1865 *Quart. J. Micr. Sci. N.S.* **3**, 99.
- Magnus, R. 1904 *Arch. Physiol.* **102**, 349.
- Matthes, K. 1930 *J. Physiol.* **70**, 338.
- McCrea, E. D. & McSwiney, B. A. 1928 *Quart. J. Exp. Physiol.* **18**, 301.
- McCrea, E. D., McSwiney, B. A. & Stopford, J. S. B. 1925 *Quart. J. Exp. Physiol.* **15**, 201.
- Oestergren, H. 1903 *Z. wiss. Mikr.* **19**, 300.
- Ogawa, F. 1934 *Sci. Rep. Tôhoku Imp. Univ. Biol.* **8**, 345.
- Ogawa, F. 1939 *Sci. Rep. Tôhoku Imp. Univ. Biol.* **13**, 395.
- Pavlov, I. P. 1910 *The Work of the Digestive Glands*, 2nd ed. London.
- Plattner, F. & Hunter, H. 1930 *Pflüg. Arch. ges. Physiol.* **225**, 19.
- Quatrefages, M. A. 1847 *Ann. Sci. Nat.* (3) **7**.
- de Ribaucourt, E. 1901 *Bull. Sci. Fr. Belg.* **35**, 211.
- Rubaschoff, S. 1912 *Internat. Beitr. Ernähr. Stor.* **3**, 462.
- Smallwood, W. M. 1923 *Proc. Nat. Acad. Sci., Wash.*, **9**, 99.
- Smallwood, W. M. 1926 *J. Comp. Neurol.* **42**, 35.
- Smirnow, A. 1894 *Anat. Anz.* **9**, 570.
- Wells, G. P. 1937a *J. Exp. Biol.* **14**, 117.
- Wells, G. P. 1937b *J. Exp. Biol.* **14**, 290.
- Wu, K. S. 1939a *J. Exp. Biol.* **16**, 184.
- Wu, K. S. 1939b *J. Exp. Biol.* **16**, 251.
- Young, J. Z. 1933 *Quart. J. Micr. Sci. N.S.* **75**, 571.

PILGRIM TRUST LECTURE

Organization of American scientists for the war

BY KARL T. COMPTON, *President, Massachusetts Institute of Technology*

(Read 20 May 1943—Received 20 May 1943)

INTRODUCTION

In ordinary times any American scientist would deem it a distinguished honour to be nominated as the Pilgrim Trust lecturer. In these extraordinary times the significance of this lectureship is enhanced by the fact that American and British scientists are working hand in hand, not only to advance science as an important aspect of man's culture, but now especially as a powerful tool for the preservation of our opportunities for continued life, liberty and pursuit of happiness. Consequently, when Sir Henry Tizard transmitted your invitation for me to deliver this lecture I accepted the invitation with profound appreciation and humility.

There were also two personal aspects of this invitation which aroused in me a sentimental reaction. The establishment of the Pilgrim Trust lectures was announced to the Royal Society by your late colleague, Sir William Bragg, in his presidential address at the anniversary meeting of the Society in 1937. Not only has Sir William Bragg, together with his distinguished son, been an inspiration to my generation of American physicists, but it happens that he delivered the address at the Graduation Exercises of the Massachusetts Institute of Technology on the occasion of my inauguration as president of that institution in 1930. I recall very vividly his remarks on that occasion and my feeling that his presence was an inspiration to me at that time when, with considerable trepidation and regret, I moved from the research laboratory into an administrative office.

In that address Sir William traced the development of modern institutions for technical instruction and drew his illustrations from the lives and work of Count Rumford, Thomas Bernard and Michael Faraday. Permit me to read several excerpts from Sir William's address.

'A school of technology is an expression of the wish to gather together the growing knowledge of nature, of natural materials and laws, and to make that knowledge of service to mankind. . . . Where and when did this movement begin? It is not of great antiquity. Perhaps the seventeenth century includes its first recognizable appearance, marked by the new-found delight in careful experimental examination of the world. The Royal Society of London was founded in the troubled times of conflict between King and Parliament: where we read the account of the early proceedings of the Society we are reminded of boys let out of school, running quickly to brook and hedgerow to examine the extraordinarily interesting things

to be found there at every turn. A very important account of the methods of ventilating mines in Belgium comes next to the careful description of some animal monstrosity such as a calf with more than the proper allowance of heads or legs. The professor of astronomy at Oxford writes on the use of incubators in Egypt. Early discussions on tide prediction, then a very important matter because the old harbours did not easily accommodate the large boats wanted for the American trade, lie side by side with a disquisition on a remarkable set of teeth and pertinent references to the properties of sugar. The hodge-podge is readily understandable. The infant was beginning to notice and ask questions. . . . We can well be amused at all this, remembering that the infant has now become a very sturdy youth: and that the proceedings of these early fellows of the Royal Society were only the first results of the new regard for natural knowledge.'

Sir William then proceeded in most interesting fashion to sketch the significant facts in the life, work and social viewpoint of Count Rumford and Thomas Bernard, and pointed out that the wealth of Bernard and the stimulation given by Benjamin Thompson, both of them men who spent their early lives within a few miles of my home in Massachusetts, led to the establishment of your Royal Institution.

Then, after drawing lessons from the life of Michael Faraday, Sir William concluded with these words:

'Science was not merely a collection of inventions to be applied by the rich for the comfort of the poor. It was a glorious purpose to be shared by all mankind. We must try to understand the world in which we live, for our own enjoyment, for the training of our minds, for the enrichment of our souls' contemplation, for the means whereby we may help each other. It is in this spirit that we may try to do our work. It is true of course that we must work for ourselves, instructors have to earn their living and students must come to learn how to earn theirs. As the world is made it must be so, but also the world is so made that the vision of its wonder and of the delight of mutual service and the happy task of exploring the one to help us in achieving the other, can light our lives for us as the sunshine lights the earth.'

As we think of the noble ideals which Count Rumford, Thomas Bernard and Michael Faraday held for science as that which 'can light our lives for us as the sunshine lights the earth', it is a grim and discouraging contrast to see the scientists of the world engaged to-day in developing new instrumentalities for destruction or other instrumentalities for protection against the destruction which would be wrought upon us by the engines of war of our enemies. The fact that this is so is a grim reminder that our skill in statesmanship and our art and ethics of Christian living have not kept pace with our ideals. We have no alternative now but to apply our knowledge of science in every aspect to serve us in our struggle for survival and to preserve for us that opportunity for which our race has struggled throughout the centuries—the opportunity to live and work in peace and freedom.

As the second personal note, let me lay claim to scientific kinship with your body as one who might be called a scientific grandchild of another one of your late

leaders, Sir Joseph Thomson. We in America affectionately called him 'J.J.', as I understand you also did in England, and we look upon him as the progenitor of that tribe of physicists who interested themselves in the conduction of electricity through gases. Taking him as the founder of that tribe, one of the second generation, your Sir Owen Richardson, was my guide and inspiration during my graduate student years at Princeton University. For his sake I am sorry that I did not turn out to be a more productive pupil, but the interests and satisfactions I have had in the field of research I owe more to him than to any other man.

As you know, we in America have two principal scientific societies which are broadly representative of all the fields of science and which are rather parallel to two of your principal scientific bodies, the Royal Society of London and the British Association for the Advancement of Science.

Dr Frank B. Jewett, President of our National Academy of Sciences, has asked me to deliver his personal message to the President of the Royal Society of London, Sir Henry Dale, and in addition he requested me to express to you the admiration which is felt by the members of the National Academy of Sciences for the magnificent manner in which the scientists of Great Britain have thrown the whole weight of their energies and abilities to master the innumerable technical problems arising in this war. He wanted me to assure you that in so far as we can do likewise, we in America are making a sincere effort to handle our similar problems and co-operatively to supplement the great work which you are doing.

Dr Isaiah Bowman, last month elected president of the American Association for the Advancement of Science, also gave me a message of greeting to British scientists from which I quote as follows:

'Now that the war has advanced to the stage at which we begin to talk of post-war plans, we feel more than ever the need for collaboration between Great Britain and the United States. While there is no such thing as an Anglo-American bloc in world politics there is such a thing as close comradeship in the fight for principles. This comradeship we feel whenever we deal with the leaders of Britain and whatever the field of interest. I venture to predict that, whatever difficulties may arise, we shall find that comradeship and agreement upon principles will ever mark our future relations. This belief is based upon our widely recognized common responsibility for the peace and safety of mankind in the years after the war. If England is being changed by the war the United States is changing just as rapidly. Once our President was able to report on "the state of the Union", as our Constitution provides, almost without touching on foreign affairs. Two world wars have changed both the tenor and the scope of such messages. The state of the Union now includes the state of the world. This conception of the state of the Union lays new obligations upon us all. The scientist can no longer report on the state of the sciences. He must report on the impact of science upon society. He must make use of the qualities of mind that science fosters in dealing rationally with the terrible waste in vital resources that war imposes upon the human species. We may hope that the day will soon come when every mature man and woman

will feel himself responsible for the state of the Union and act responsibly in that sector of our common life committed to his care, no matter how small the sector may be, no matter how humble.

'We say these things while recalling again how great an inheritance they represent from the unfolding political life of Great Britain. No one can speak of liberty and political responsibility and community enterprise without echoing truths that were discovered by centuries of experiment and experience on the part of rulers and ruled in the English political system. Thus, no matter where one starts in estimating future problems and future responsibilities, one ends by recognizing the special bond between America and Britain, by acknowledging the rich inheritance that has been responsible for so many strong elements in American life, and by elevating the comradeship that we both feel and need.'

And now I come to the subject of my address. I have chosen to speak on a subject pertinent to the war, to describe to you the manner in which American scientists have organized to make their contribution to the same cause which has mobilized your efforts. In so doing, I trust that I shall be within the spirit, if not the letter, of Sir William's directive for these lectures. For he said, 'Such lectures would associate workers in a common task', and surely our common task right now is to direct our scientific resources for victory.

In these days, when numbers of scientists are crossing the Atlantic in both directions on special missions, a better understanding of each other's organizations may be practically helpful. For I frankly confess sympathy with one of your number who recently told me that he found the American organization of inter-related scientific groups a bit complicated. I can only draw cold comfort from the fact that, complicated as it is, the scientific organization is far simpler than that of our governmental departments and bureaux generally. But that is another story.

PEACE-TIME ORGANIZATION OF SCIENTISTS IN THE U.S.

Let me first give an over-all picture of the scientific and technical organizations of the United States as they exist in peace-time. After this brief review I shall pass to a discussion of the special scientific organizations for war, which is the subject of more particular interest to us at this time.

The scientific and engineering work in the United States may be discussed under three categories: first, the agencies of the Federal Government, exclusive of the Armed Services; secondly, the agencies within the Armed Services; and thirdly, the non-governmental agencies.

Federal Bureaux. The scientific services of the Federal Government in peace-time are spread through about forty federal bureaux, of which eighteen can be called primarily scientific. Their operations involve only about half of 1% of the total peace-time federal budget, but their work is absolutely essential to the national welfare in agriculture, manufacture, commerce, health and safety. The personnel of all these bureaux operates under the Civil Service.

From the point of view of size of personnel and budget, the scientific services under the Department of Agriculture stand first in the list. Probably these scientific establishments, however, are not as well known generally as those of some of the other departments because their research work is quite largely spread through a great number of agricultural experimental stations distributed throughout the various states of the Union and operated co-operatively between the Federal Government and the States. Most of the bureaux in Washington are primarily of an administrative character, but there are several which also conduct centralized research, as, for example, the Bureau of Chemistry and Soils and the Food and Drug Administration. Until recently the U.S. Weather Bureau operated under the Department of Agriculture, but a few years ago it was transferred to the Department of Commerce, largely because the requirements of air transportation had taken the lead in demanding more accurate and refined methods of weather forecasting than those which had served reasonably well in the past to provide for the needs of agriculture.

Many of you will probably recognize some of the more important of these governmental scientific bureaux, as, for example, the National Bureau of Standards under the Department of Commerce, the Geological Survey, the Bureau of Mines, the Bureau of Mineral Statistics and Economics under the Department of the Interior, and the National Institute of Health under the U.S. Public Health Service.

Of particular interest because of its unique character is the National Advisory Committee for Aeronautics which was established during the last war and which operates three great research establishments. Until recently the work of the N.A.C.A. was centred in the aerodynamical research programme at Langley Field, Virginia. Several years ago there was added another aerodynamical research establishment named the 'Ames Laboratory' at Moffett Field, in California, and quite recently still another large research and development establishment for aircraft engines in Cleveland, Ohio. Also, under the N.A.C.A., there are currently some eighty research projects being carried on at universities under contract. This obviates unnecessary duplication of facilities in a government laboratory and maintains a group of university scientists and engineers in close contact with the problems of aeronautical research.

The administration of this organization is also unique among our federal scientific agencies in that its controlling body is a committee which serves without salary and has been composed of men of such high character and distinction as to render it completely free from political influence. This committee is provided with representation from the most interested branches of the Army, Navy and Governmental Departments, but the chairman and the majority control reside in a body of citizen scientists appointed by the President who, in practice, has followed the recommendations of the chairman in appointments to fill vacancies. The present chairman of the N.A.C.A. is Professor J. C. Hunsaker, head of the Departments of Mechanical and Aeronautical Engineering at the Massachusetts Institute of

Technology, and incidentally, while a very young man, the designer of the first American airplane to fly the Atlantic.

Turning now to the United States Armed Services, I can best describe their research and development work as principally a co-operative effort between the services themselves and American industrial companies, with occasional participation from the research laboratories of the technological and educational institutions.

Army. Each branch of the Army contains a technical division under which operate laboratories or arsenals in which a certain amount of research and development work is carried on, but whose activities consist for the most part of testing and proving new war materials or equipment. Thus the technical staffs in the various branches of the Army and Navy have the threefold duty of planning and co-ordinating an extensive programme of research and development carried on in the industrial laboratories, of organizing and conducting research programmes in their own establishments, and of carrying on the extensive operations of proving and testing which result in the acceptance of new devices and the drafting of specifications for production orders.

Among the principal Army establishments in which such work is centred, I would mention particularly those falling under the Ordnance Department, the Signal Corps, the Chemical Warfare Service, and the Army Air Forces. The Ordnance Department operates a great proving ground at Aberdeen, at which is centred most of the proving and testing of ordnance and research on ballistics for arms of all types. In addition it operates five principal arsenals. The Watertown Arsenal is concerned principally with the manufacture of mounts for large-calibre guns and is the principal centre for research and technical service in the field of metallurgy. The Picatinny Arsenal is devoted to the testing of explosives and the design and operation of pilot plants as guides to the industrial producers. The Rock Island Arsenal carries on research and development in the field of oils and lubricants. The Frankford Arsenal supplements the Aberdeen Proving Ground as a testing and a development centre for small arms. The Tank Arsenal in Detroit is the centre for the design and testing of tanks.

In the Signal Corps the technical division is divided into three principal branches: the Ground Signal Branch, the Electronics Branch and the Aircraft Radio Branch. The research, development and testing work carried on under the Signal Corps is divided principally between the signal laboratories at Fort Monmouth, Camp Evans, Camp Coles, Eatontown, and Toms River. The Signal Corps also maintains a large co-operative establishment working with the Army Air Forces at its principal centre, Wright Field.

Until quite recently the research and development work of the Chemical Warfare Service was centred in its great Edgewood Arsenal. As the threat of war came closer, however, a few years ago, and since a very large portion of the facilities at the Edgewood Arsenal are taken up by production, the Chemical Warfare Service established a subsidiary research laboratory and took over for this purpose the

newly erected Chemical Engineering Laboratory of the Massachusetts Institute of Technology.

Despite the great expansion of the Army Air Forces, this service has continued to concentrate its research, development and testing activities at its huge establishment at Wright Field in Ohio. There are of course many other centres at which extended service testing goes on, or at which new equipment is installed in aircraft, but Wright Field remains the headquarters for the research and development work of the Army Air Forces.

The co-ordination among all these various technical services of the Army is maintained by two types of agency. Within each branch of the Army is a Board which has general supervision over technical matters within that branch. Examples are the Coast Artillery Board and the Army Engineers Board. When a project has been approved by one of these boards it is next discussed by the appropriate Technical Committee, composed of members of this branch of the service, and other branches which may be concerned with the project. If this Technical Committee also approves the project it goes as a recommendation to the general staff which presumably issues the appropriate directive.

Mention should be made also of the Army Medical Corps, within which a significant amount of research is conducted under the general supervision of the Surgeon-General of the Army.

Navy. The Naval Observatory and the Hydrographic Office, which are under the Chief of Naval Operations, have obvious functions in research and development work. The Marine Corps does some research, but naturally depends to a large extent on the Army and the various Bureaux of the Navy. All the Bureaux of the Navy Department, Ships, Ordnance, Aeronautics, Naval Personnel, Supplies and Accounts, Medicine and Surgery, Yards and Docks, do research and development work, though naturally the matériel Bureaux conduct the greatest volume.

All research work of the Navy Department is tied together through the office of the Co-ordinator of Research and Development, which office also arranges co-ordination with the Army, with other government departments, and with the numerous civilian agencies which I will mention later.

Under the Bureau of Ships the Naval Research Laboratory near Washington is a centre for all matters of fundamental research including radio, electronics, chemical warfare defence, etc. The David W. Taylor Model Basin, also near Washington, is the primary station for research and development of ship structures, propeller and hull design. The Naval Boiler and Turbine Laboratory at Philadelphia is concerned with all matters of boiler research, testing and design including fuel, composition, quality and nature of boiler fuels, ceramics, etc., and also for research, test and development of main propulsion turbines. The U.S. Naval Engineering Experiment Station at Annapolis, Md., is assigned all problems of research, test and development of mechanical equipment in ships other than main propulsion, and it also has a well-equipped Diesel-engine laboratory. The principal metallurgical laboratory for the Bureau of Ships is also located at the Engineering

Experiment Station. In New York there is located the Materials Test Laboratory which handles all matters of research, test and development of electrical materials and equipment, acoustical equipment, optical and navigational material and equipment, plastics and allied materials. There are Rubber and Paint Laboratories at Mare Island, California, and an inspection test laboratory in Pittsburgh, Pa., where line-production methods for chemical analyses are set up which permit a capacity of about 5000 chemical analyses per week with a minimum of personnel and equipment. In addition to the above each Navy Yard is equipped with an industrial laboratory to serve the purposes of the Yards. It has been found possible to place specialized problems in some of these laboratories such as the development of chain and rope in the Boston Navy Yard. The assignment and progress, as well as general administration of all research, development, and test work, is carried out by the Bureau of Ships in Washington in order most fully and effectively to co-ordinate all work and to collect, apply, and distribute the results.

Under the Bureau of Ordnance there are the Dahlgren Proving Ground, the Naval Gun Factory at Washington, whose research department includes the Naval Ordnance Laboratory, the Naval Powder Factory near Washington, and the Newport Torpedo Station. In addition there are establishments devoted to mines, counter-mines, nets and the like.

Under the Bureau of Aeronautics there is the Naval Aircraft Factory in Philadelphia, the Cedar Point Flight Testing Field near Washington, and the Aircraft Armament Laboratory and Testing Field at Hampton Roads.

Research in medicine and surgery is directed by the Research Division of the Bureau of Medicine and Surgery, using many facilities but largely those of the U.S. Naval Medical Research Institute at Bethesda near Washington, and the Medical Research Laboratories at Pensacola and New London.

Civilian agencies. I pass now to the non-governmental scientific organizations in the United States, most of whose members are attached to the staffs of some 600 colleges, universities and engineering schools, some 2000 industrial research laboratories, and other specialized research institutes. Do not be alarmed when I begin by saying that these comprise well over one hundred nationally recognized scientific and engineering societies, exclusive of the social sciences. Of these, only a few are general in scope in the sense that they cover broadly the entire field of science. Largest of these is the American Association for the Advancement of Science, a close parallel to your British Association, with a direct membership of about 24,000 and an indirect aggregate membership of about a million through the 187 associated and affiliated societies. Of a more exclusive character and without the affiliated and associated societies are the American Philosophical Society and the American Academy of Arts and Sciences.

Unique among the scientific organizations of the United States is the National Academy of Sciences. In March 1863, during a crisis of our Civil War, Congress established the National Academy of Sciences, and President Lincoln signed the Act of Incorporation. This Act specified that 'the Academy shall, whenever called

upon by any department of the governments, investigate, examine, experiment and report upon any subject of science or art, the actual expense of such investigations, examinations, experiments and reports to be paid from appropriations which may be made for the purpose'. There was also the provision in the charter that, except for the actual expenses of these activities, neither the Academy nor any member of the Academy is entitled to receive any compensation whatsoever for such services. Although the membership is legally limited to 450, the actual membership in the Academy has never exceeded its present enrolment of 350.

Outside of its services in war-times, perhaps the most noteworthy public service by the Academy was its geological and engineering investigation of the slides which at one time threatened to prevent the successful consummation of the Panama Canal. However, the utilization of the Academy by the government has been rather 'spotty'. Under some administrations the Academy has been used rather extensively, and in other administrations has been more or less forgotten by the government. In this respect I believe that your Royal Society has had a more consistent role of usefulness.

One inevitable characteristic of this type of organization, in which membership is considered to be the highest scientific honour of the country, is that membership, like scientific recognition, is likely to come to a man after he has passed the peak of activity in his scientific career. For this reason the Academy has been able to perform an excellent function of the 'scientific elder statesmen' variety. It has zealously kept itself free from all types of political influence. Its ideals have been unselfish service, integrity and scientific competence. Frequently, however, probably in the great majority of cases, when a very active research programme has to be undertaken, many of the personnel best adapted for the particular job are not found within the membership of the Academy.

During the last world war in Europe, but before the United States had become a participant, President Wilson by executive order requested the National Academy of Sciences to establish the National Research Council as a measure of national preparedness. This organization operated so usefully during the war that after its termination, in April 1919, the National Research Council was perpetuated by the National Academy of Sciences at the express request of President Wilson.

This National Research Council is organized into nine permanent divisions covering the various fields of scientific research and of scientific administration. These divisions are composed of appointed members and also of representatives from many of the scientific and engineering societies and branches of the government. Because of this wide representation the National Research Council is a most effective agency for finding just the right persons to do any specific scientific job.

During the present war the National Academy and the National Research Council have been called upon to perform many important services, some of an advisory character and some involving the placing of contracts for research and development work in various laboratories.

Among the nearly 200 committees operating under the National Research Council, the following are typical of those concerned with the war: Aviation Medicine, War Metallurgy, Passive Protection against Bombing, War Use of Research Facilities, Tin Smelting and Reclamation, Clothing, Shock and Transfusions, Treatment of Gas Casualties, War-time Diet, and Selection and Training of Service Personnel.

WAR-TIME SCIENTIFIC ORGANIZATION

In spite of the apparently complete peace-time organization which I have just described, it has always been our experience, in the time of great emergency, that it appears advisable to establish temporary new agencies to deal particularly with the emergency. For example, I happened to be attached to one of these temporary agencies during the last war, and I mention the matter not only by way of illustration but also because it will enable me to relate an anecdote about your late distinguished colleague, then Sir Ernest Rutherford.

This agency was the Research Information Service, set up jointly by our Military Intelligence, Naval Intelligence and Council of National Defence, with offices in Washington, London, Paris and Rome. The function of these offices was essentially the same as that of the Scientific Liaison Offices which have been operating so effectively between units of the British Commonwealth and the United States during the present war.

The head of the Research Information Service in London was the late Professor Bumstead whom some of you doubtless remember. I was attached to the Paris office and happened to be temporarily in charge during the time when an allied conference on submarine detection was arranged in Paris under the auspices of this office.

One of the delegates from Great Britain was Sir Ernest Rutherford who had been collaborating closely with the French physicist, Paul Langevin, in the development of underwater supersonic devices. The day before the conference, when the British and American delegations came over from London, Rutherford was not present but he sent me a letter, delivered by Professor Bumstead, stating that some very recent experiments which he and his research assistant had been carrying on in the Cavendish Laboratory had apparently indicated success in disintegrating the nucleus of the hydrogen atom. 'If this is true', Rutherford wrote, 'it is a fact of far greater importance than the war.' He went on to say that he was in the midst of a second experiment to check these startling findings and that he would be delayed a couple of days pending the termination of this experiment. Then Rutherford added as a postscript: 'Tell nobody about this because I may be mistaken.' Later it developed that what Rutherford had actually done had not been to disintegrate the hydrogen nucleus, but rather to disintegrate the nuclei of nitrogen atoms. So far as I know, Rutherford's letter to me was the first written indication of success in the long, long struggle to produce by artificial means a transmutation of one chemical element into another. I wish I had kept that letter

and had turned it over to Professor Eve at the time when he was writing his interesting biography of Ernest Rutherford. But to return to our topic.

I have frequently tried to analyse the reasons for the establishment of special scientific agencies during times of crisis. The reasons I think are varied and rather fundamental. One of them is that every great crisis involves conditions so different from the normal situation that the types of organizations which can survive and operate during peace-time are not adequate to meet the emergency. It may be, for example, that the emergency calls for exercise of very extensive administrative functions, such as the supervision of research projects and the disbursement of large governmental funds to a far greater extent than in peace-time. Hence a peace-time body of scientists organized primarily to exercise advisory functions may not be organized in a manner suited to prompt and efficient executive action. Another reason is the impossibility of always maintaining in the administrative positions of peace-time agencies the personnel who would be most effective for handling important projects in a war emergency. Men who have the proper capabilities are frequently too busy and too active in other directions to be willing to hold down positions in a peace-time organization which is relatively inactive. Consequently when the emergency comes, the only alternatives may be to change the leadership in the existing organizations, a difficult if not impossible process, or to set up new temporary agencies to deal with the emergency.

Whatever the reasons may be, this present war emergency has run true to form and has resulted in the establishment of a group of special agencies of temporary character which I shall proceed now to describe. It is these agencies which are carrying the principal burden of the scientific research and development work related to the war, in the United States.

The National Roster of Scientific and Specialized Personnel was established early in July 1940, when President Roosevelt approved a project for making available in one central office an index of all American citizens who have special scientific or professional skill. Headed by President Leonard Carmichael of Tufts College, this agency operates under the War Man-power Commission under the Office for Emergency Management of the executive office of the President. As a result of information secured from questionnaires sent to all members of all scientific and professional organizations in the country, and supplemented by other information, an elaborate punch-card system has been set up in which practically every person in the country with specialized training or skill is listed with reference to his or her major professional fields and with the addition of a great deal of supplementary information regarding special interests, languages read or spoken, foreign countries travelled in, previous experience in the armed services or in industry, etc.

There are altogether fifty-nine special fields listed in the roster, falling under the general categories of administration and management, agricultural and biological sciences, engineering and related fields, humanities, medical sciences and related fields, physical sciences, and social sciences. At the present time the total number

of persons in this roster is about 600,000, including, in October last, 71,511 chemists, 7297 mathematicians, 10,080 physicists or astronomers, 4559 radio engineers, 14,729 electrical engineers, only 408 professional philosophers, and the smallest entry in the list is 142 speleologists.

As an illustration of the manner of use of this roster I quote the following paragraph from a report by Dr Carmichael:

'How would you like to get an order for the names of all Americans who possess a knowledge of epidemiology and chemotherapy, who are competent in the diagnosis and control of *Endamoeba histolytica*, and other protozoan infections, have a knowledge of the Hindustani language, are skilled in the operation and use of specialized bacteriological research apparatus and who have travelled in the tropics?' To secure this information the stops of the punch-card sorting mechanism are pushed in at the appropriate places, the cards are ground through the machine, and all of those which conform to the above specifications fall out together.

The roster was originally conceived to serve governmental agencies who might request information on scientific personnel. More recently, as serious man-power shortages have developed both in industry and in education, and as the armed services have become more and more concerned over the most effective use of all scientifically trained personnel, the roster has been used to an increasing degree in connexion with placement work and to give the supply and shortage data on professional groups. Up to the middle of last month the National Roster had certified more than 140,000 names of specialists to various agencies engaged in the war programme in the United States.

Office of Scientific Research and Development (O.S.R.D.). Most important of the scientific agencies established specially to deal with problems of this war is the Office of Scientific Research and Development, whose Director is Dr Vannevar Bush, President of the Carnegie Institution of Washington. It was created by executive order of the President in June 1941, and under it operate the National Defence Research Committee, which had been established just a year earlier, and also the more recently established Committee on Medical Research. The O.S.R.D. is directed to co-ordinate, and where necessary supplement, the scientific research and development work relating to the war among civilian agencies as well as those of the government, including the Armed Services. To facilitate this co-ordination the advisory council to the Director of O.S.R.D. includes high-ranking representatives from the War and Navy Departments, the Chairmen of the National Advisory Committee for Aeronautics, the National Defence Research Committee and the Committee on Medical Research, and, by invitation, the President of the National Academy of Sciences and the Director of the newly established Office of Production Research and Development of the War Production Board.

The principal research and development activities of the O.S.R.D. are carried on under contracts with appropriate research institutions, these contracts being financed out of an annual Congressional appropriation. At the present time these contracts involve expenditures at the rate of about \$100,000,000 per year, and

there are currently active about 1400 contracts with about 200 industrial laboratories and 100 educational or special research institutions. About 6000 scientists and engineers of professional grade are engaged on these contracts, with the assistance of a considerably larger number of technicians of various types.

To facilitate interchange of information between O.S.R.D. and our British colleagues, an O.S.R.D. Liaison Office was established with offices in Washington and London, now headed by Dr Caryl P. Haskins and Mr Bennett Archambault, respectively. These, in co-operation with the similar liaison services of Great Britain, Canada and, less extensively, Australia and South Africa, have served well to knit together our joint scientific efforts.

The National Defence Research Committee (N.D.R.C.) operates to recommend to the Director of O.S.R.D. research and development contracts in the field of instrumentalities, devices and mechanisms of warfare. Under the chairmanship of President James B. Conant of Harvard University, this Committee is composed of four civilian scientists plus one representative each from the Army and Navy, and the Commissioner of Patents. Feeding into it come the recommendations from nineteen divisions, most of which are subdivided into several sections. These divisions and sections are each built around a specific functional concept such as fire control or subsurface warfare or explosives. However, there are two divisions which are in the nature of 'catch-alls'. For example, the Division of Physics and the Division of Chemistry can be defined as handling everything in these respective fields which does not fall under any one of the more sharply defined divisions.

In addition to the nineteen divisions of N.D.R.C. there are two panels concerned respectively with applied mathematics and engineering. The difference between a division and a panel is suggested by the fact that the Fire Control Division, for example, is concerned with the development of fire control instruments, whereas the Applied Mathematics Panel is not concerned with the development of applied mathematics as such, but rather with the use of mathematics to aid in accomplishing the objectives of the various divisions. For this reason the Applied Mathematics Panel includes membership on each divisional committee in which applied mathematics is likely to be important. The Engineering Panel serves all the divisions to expedite the transition from the stage of research and development to the stage of quantity production under Army or Navy contract.

Intimate contact between N.D.R.C. and its divisions on the one hand, and the Armed Services on the other, is maintained at several levels by an extensive organization of Army and Navy liaison officers who have proved invaluable as channels for acquainting N.D.R.C. with the needs and desires of the Armed Services for new equipment and for making arrangements for demonstrations and service tests.

Proposals for research or development projects come to N.D.R.C. from a wide variety of sources—requests or suggestions from the Army or Navy, proposals from industrial or academic research laboratories, promising inventions transmitted to N.D.R.C. from the National Inventors' Council, or in many cases

projects originating within the N.D.R.C. committees themselves. However, the N.D.R.C. has complete freedom in making its decisions on the projects which it recommends to the Director of O.S.R.D. and the priority attached to these projects, and the Director of the O.S.R.D. has complete freedom in his own judgement to authorize the recommended contracts.

For reasons of security no person serves as a member of any N.D.R.C. committee unless he has been 'cleared' by the Army and Navy Intelligence Offices, after investigation. Similarly, all personnel of the contractors working on the research and development projects are 'cleared' by these intelligence offices to whatever degree is deemed advisable in virtue of the degree of secrecy attached to the project.

The Committee on Medical Research (C.M.R.), under the chairmanship of Dr A. Newton Richards of the Medical School of the University of Pennsylvania, is in every respect parallel to the National Defence Research Committee in its organization and methods of operation. It deals exclusively with problems of war medicine, such as shock, immunization or protection against types of diseases characteristic of the present theatres of war, etc. Though considerably younger and smaller than N.D.R.C. in both personnel and budget, it already has a record of substantial accomplishment.

Joint Committee on New Weapons and Equipment (J.N.W.). The organizations described thus far have proven effective in organizing and administering research projects and in maintaining close relationships and exchange of information with the Armed Services and our British allies. In respect to the Armed Services, however, these relationships are primarily at the research and development level and for a time lacked one very important element necessary to make the work fully effective in the war. This missing element was an intimate relationship between the research and development agencies and the Highest Command of the Army and Navy who have the responsibility of planning the military or naval operations in which newly developed weapons might be used effectively or for which new devices should be developed. In order to fill this gap the U.S. Joint Chiefs of Staff in May 1942 established the Joint Committee on New Weapons and Equipment, composed of Dr Bush, Director of O.S.R.D., as Chairman, the Assistant Chief of Staff G4 of the Army (now Brigadier-General Moses) and the Chief of the Readiness Division of the Navy (now Rear-Admiral De Laney).

J.N.W. is charged by the Chiefs of Staff with correlating the research programme of army, navy and civilian agencies. It acts through subordinate bodies of which the special mission in which I am at present engaged in England is an example.

Through J.N.W. any new weapon whose potentialities appear to be unusually significant is brought directly to the attention of the High Command for their consideration in the planning of future operations. Conversely, J.N.W. offers a direct channel through which the High Command can pass down to the research scientists a request for development of any particular instrumentality which could be particularly effective in connexion with some contemplated operation. This

type of liaison between the scientists and the High Command is new in the United States. Its possibilities are still being explored and developed, but it can be said definitely that it has already demonstrated its possibilities of great value in the war. It is a move in a desirable direction in which you have gone further than we have gone.

National Inventors' Council. War is a great stimulus to invention, not only in the research laboratories of a country, but on the part of great numbers of its citizens, some of whom are technically competent and most of whom are uninformed but sincere in their desire to be helpful. Any actively operating research organization like the O.S.R.D. or the Naval Research Laboratory could be quickly bogged down under the deluge of ideas and inventions induced from all sources by the war. It is very important for purposes of morale that these inventors and would-be inventors be sympathetically handled. It is also important that their ideas be expertly examined to make sure that really worth-while ideas are not brushed aside, even though experience has shown that perhaps only one in one hundred thousand is new and significant.

To give such sympathetic and expert consideration and to screen the interesting suggestions out of the great mass, the National Inventors' Council was established in June 1940, in close association with the U.S. Patent Office in the Department of Commerce, under the chairmanship of Dr Charles F. Kettering, Vice-President in charge of research for the General Motors Corporation. All suggested inventions relating to the war from any source and submitted to any agency or person in the government are channelled through this National Inventors' Council (unless they happen to come initially to an appropriate agency which is immediately interested in pursuing the matter). They pass through the hands of an expert staff of examiners who select those inventions which appear to have merit and bring them to the attention of the appropriate agency.

Office of Production Research and Development of the War Production Board. Until recently the organized war research efforts in the United States failed to include the very important category of research aimed at the development of substitute materials in fields where shortages exist, or of improved methods of production and manufacture. It was apparently assumed that the commercial interest of the production companies would lead them automatically to take care of this situation. However, under the pressure of war-production orders, limitations of man-power and materials, and financial regulations, the normal peace-time incentives to such research and development work by companies proved inadequate to meet the needs of the situation. Consequently, last September, there was established in the War Production Board an Office of Production Research and Development under the directorship of Dr Harvey N. Davis, President of the Stevens Institute of Technology. This agency is still in the process of organization to operate somewhat along the lines of the Office of Scientific Research and Development but with primary responsibility for materials and method of production rather than for devices and instrumentalities of warfare. It is regrettable

that we did not have the foresight to establish this much-needed agency at a much earlier date, but it has already begun its operations and we hope that it may be enabled to play an important role during the balance of the war.

Engineering, Science and Management War Training Programme. Though not directly concerned with scientific research, a review of the scientific war agencies in the United States would not be complete without at least a brief reference to the efforts to increase the supply of technically trained personnel to meet the increasing demand for such personnel in every field of war activity. In October 1940 a special engineering training programme was organized under the U.S. Office of Education and financed by Congressional appropriation. Later, this programme was extended to include also training in science and industrial management. It operates at both the collegiate and the technical school levels, and its magnitude may be appreciated by the fact that, even in its first year of operation, it put through its specialized courses approximately ten times as many students as graduated in that year from the regularly established engineering colleges. Most but not all of the work was carried on in night schools, and the whole programme has been decidedly helpful in relieving the technical man-power shortage.

Army and Navy Technical Training Programmes. At the present time the Army and Navy are jointly establishing a very extensive programme for the training of their own younger personnel in such fields as aeronautical engineering, naval architecture, electronics, communications, automotive engineering, etc., through contractual arrangements with several hundred of the nation's colleges and universities. Under these programmes it is anticipated that approximately 250,000 selected young men in uniform will be detailed for this training at educational institutions during the coming year, the duration of such training to vary from field to field and individual to individual, in accordance with the needs of the situation and the performance of the individual. These special collegiate programmes are intended to supplement, at the higher level, the very much larger technical training programmes which the Army and Navy are conducting in their own establishments.

CONCLUSION

I conclude this factual, over-long, but I hope usefully informative address on a note of faith and optimism which I am sure is shared by the allied scientists on both sides of the Atlantic. Each of us concerned with some phase of the war effort is aware of some very significant new applications of scientific research in the war. For most of us, this knowledge is largely restricted to the special fields in which we ourselves have been working. Of necessity, the general public knows only in a vague way about some of these things and nothing at all about most of them.

When victory has been won, and the whole story of these scientific accomplishments can be told, it will indeed be a thrillingly interesting recital. Out of it all will come, not only its important contribution to victory, but a number of exceedingly significant results of permanent peace-time value. It is already evident that

many of these war-time developments will have very useful peace-time applications, whose contributions to our standards of living and general prosperity and comfort will help to compensate for the ravages wrought by the war. Scientists will have a renewed faith in the worth-whileness of their work, and will continue their intellectual and practical endeavours with the increased power that has come from the experience of 'team-work' on war problems. The general public, and especially the governmental and industrial leaders, will have greater appreciation of the value of science and scientists, both pure and applied—and this should result in permanently increased support of scientific research in the universities, industries and governmental agencies. These, I trust, will be some of the long-term gains to which we may look forward as the result of the temporary concentration upon practical problems of survival and victory which the war has forced upon us.

With these words of optimism, I close with the hope that the next American Pilgrim Trust lecturer to address you may not feel obliged to discuss the war, but will be able to treat of some interesting aspect of the progress of science in accord with the original conception of Sir William Bragg and as a happy feature in the post-war forward march of Science.

The fate of haemoglobin in *Rhodnius prolixus* (Hemiptera) and other blood-sucking arthropods

BY V. B. WIGGLESWORTH, F.R.S.

London School of Hygiene and Tropical Medicine

(Received 2 April 1943)

In *Rhodnius* most of the haemoglobin ingested is broken down in the lumen of the gut to protohaematin which is excreted unchanged. But a small amount is absorbed and circulates in the haemolymph as kathaemoglobin (parahaematin). This is taken up by the salivary glands where it appears as a cherry-red pigment with properties similar to haemalbumin. Blood pigment is also transferred to the yolk of the eggs and becomes concentrated in the stomach of the newly hatched nymph as a bright red fluid (parahaematin). In the next few days most of this is digested in the gut to give protohaematin; some is transferred to the salivary glands to give rise to their usual pigment.

Blood pigment in the haemolymph of *Rhodnius* is taken up also by the pericardial nephrocytes and by the epithelial cells of the stomach and intestine. Here it is converted to a brown pigment (a modified haem pigment), to a green pigment (probably of the verdohaem type and resembling choleglobin), and finally to biliverdin. Biliverdin accumulates throughout life in the pericardial cells, which become bright green. In the gut it is discharged to the lumen and appears in the faeces. The free iron accumulates throughout life in the cells of the stomach and intestine; in old insects these are heavily laden with iron deposits.

After the injection of haemoglobin into the haemolymph all the above processes are exaggerated. In addition, some breakdown of blood pigment takes place in the Malpighian tubes; and the lumen of the tubes may become charged with massed droplets of biliverdin displaced from the pericardial cells. The process of excretion of this pigment by the cells is described.

Comparative studies have been made on some other blood-sucking arthropods. Hemiptera: *Triatoma infestans*, *T. brasiliensis*, *Eutriatoma sordida*, *Cimex lectularius*. Anoplura: *Pediculus humanus*. Siphonaptera: *Nosopsyllus (Ceratophyllus) fasciatus*. Diptera: *Anopheles maculipennis*, *Aedes aegypti*. Acarina: *Ornithodoros moubata*, *O. delanoei*, *Ixodes ricinus*.

In all these haemoglobin is digested in the gut more or less completely to protohaematin. It is demonstrable in the haemolymph in *Ornithodoros*, probably as alkaline haematin, and in *Ixodes*, in a form resembling methaemalbumin. It appears in modified form in the salivary glands in *Cimex* only. It is transferred to the eggs in *Cimex*, *Ornithodoros* and *Ixodes* as alkaline haematin; in *Pediculus* apparently as unchanged oxyhaemoglobin. It is broken down to biliverdin in the Malpighian tubes of *Triatoma infestans*, *T. brasiliensis* and *Eutriatoma*; in the pericardial nephrocytes of *Triatoma brasiliensis*, *Eutriatoma* (probably associated with bilirubin), and *Pediculus*; and in the lumen of the gut in *Pediculus*. An altered haematin and perhaps a verdohaem pigment similar to choleglobin have been demonstrated as intermediates in *Triatoma brasiliensis*.

It is well known that in mammals a certain amount of the protein in the food is absorbed unchanged (Verzár & McDougall 1936). To what extent this happens in insects is unknown. But the blood-sucking insects, in which a large part of the ingested protein consists of the easily recognizable haemoglobin, provide particularly favourable material both for the study of this question and for following the subsequent fate of the protein so absorbed. In the present work the blood-sucking bug *Rhodnius prolixus* has been chiefly used, but preliminary observations have been made on some other blood-sucking arthropods.

DIGESTION OF HAEMOGLOBIN IN THE GUT OF *RHODNIUS*

Figure 1 shows the anatomy of those organs in *Rhodnius* with which we shall be concerned. The adult bug is about 2 cm. in length and weighs about 80 mg. At a single feed it may ingest more than twice its own weight of blood. The blood is stored and concentrated in the capacious 'stomach', the first segment of the mid-gut. Here the haemoglobin may remain undigested for several weeks.

If the brownish red residue still present in the stomach of an adult *Rhodnius* soon after moulting (that is, a month or more after feeding on the rabbit in the preceding instar) is examined spectroscopically,* it still shows the bands of oxyhaemoglobin at 577 and 540 m μ and a strong methaemoglobin band at 630 m μ . On the addition of sodium hydrosulphite the methaemoglobin band disappears, the bands of oxyhaemoglobin are replaced by an intense single band of reduced haemoglobin, and there is now perceptible a very faint band of acid haematin at about 640 m μ , showing that even after storage for this length of time in the stomach digestion of the haemoglobin has scarcely begun.

As the blood is passed on gradually to the coiled 'intestine', it is soon digested and darkens progressively until the hind end of the gut contains only a black residue. This is finally discharged to the rectum. If the contents of successive segments are diluted with distilled water and examined spectroscopically, it is

* The spectroscopic examinations throughout this work, unless otherwise stated, have been made on solutions in glass tubing of 0.75–3 mm. internal diameter, using a Zeiss micro-spectroscope.

possible to detect the bands of oxyhaemoglobin only at the commencement of the intestine where it joins the stomach. In addition there is an intense absorption band of acid haematin with its centre at about 640 or 645 $m\mu$ and a diffuse band in the green at about 550 $m\mu$. These bands of acid haematin alone are visible throughout the remainder of the intestine. They are unaffected by sodium hydrosulphite.

The black residue in the rectum probably consists mainly of free haematin (ferric iron porphyrin) deprived of its protein moiety in the course of digestion. Dissolved in 30 % pyridine in water it shows at first the absorption spectrum of acid haematin; but gradually the solution becomes red, these bands fade, and the absorption spectrum of ferric haemochromogen (pyridine parahaematin) appears. This consists of two diffuse bands at about 565 and 530 $m\mu$, the second of these being broader and stronger. On reduction with sodium hydrosulphite the absorption bands of pyridine protohaemochromogen (ferro-haemochromogen) appear: an intense narrow α -band at about 557 $m\mu$ and a diffuse β -band at about 526 $m\mu$. Pink crystals of pyridine haemochromogen form within a few seconds if Takayama's reagent* is added to the dried residue from the gut.

If the undigested blood in the stomach or the black residue in the rectum are dissolved in concentrated sulphuric acid, the resulting porphyrins show identical absorption spectra: a narrow band with centre at 596 $m\mu$, a very faint narrow band at about 576 $m\mu$ and an intense broad band at about 552 $m\mu$. These values agree fairly well with those given by Fischer & Treibs (1926) for haematoporphyrin. If the porphyrin is precipitated in water and redissolved in 4 % sodium hydroxide in 50 % alcohol containing 15 % glycerol, it shows a four-banded spectrum as follows: 620, 569, 538 and 505 $m\mu$.

The foregoing represents the fate of the bulk of the haemoglobin ingested. While most of the globin has been assimilated, the iron porphyrin is excreted unchanged. The excreta, made up of the residue from the intestine plus the urine from the Malpighian tubes, contains no free iron (Wigglesworth 1931).

BLOOD PIGMENT IN THE HAEMOLYMPH

A certain amount of haemoglobin is, however, absorbed without being digested. The freshly drawn haemolymph, collected in a capillary tube from an amputated leg stump, ranges in tint from almost colourless to a distinct pinkish yellow. Examined spectroscopically it shows two very hazy absorption bands with their centres at about 567 and 531 $m\mu$, the β -band being much broader and slightly darker than the α -band. This spectrum is unaffected by potassium ferricyanide. On adding sodium hydrosulphite the absorption bands are immediately replaced by those of haemochromogen (558 and 527 $m\mu$), but revert to their original position on exposure to the air.

* 10 % sodium hydroxide 3 ml.; pyridine 3 ml.; saturated solution of glucose 3 ml.; distilled water 7 ml.

It is clear that the pigment circulating in the body cavity has become greatly changed. Its properties are those of 'parahaematin' or 'kathaemoglobin' (Keilin 1926). Evidence that this change takes place in the haemolymph will be given later (p. 325).

In order to determine whether the appearance of blood pigment in the body fluid is dependent on the excessive stretching of the stomach (it is not uncommon for the stomach, or the whole insect, to rupture spontaneously during feeding), six adult female *Rhodnius* were allowed to take very full feeds, and six others very small feeds, which caused no distension of the stomach or body wall. Haemolymph was collected from the six pairs of insects three days later, introduced with segments of capillary tubing of uniform bore and the concentration of pigment compared. There was much individual variation; but in four out of the six pairs the insect that had taken the very small feed had more pigment in the body fluid. The absorption of haemoglobin is clearly not dependent on excessive distension. This conclusion is supported by observations on the first stage nymph (p. 329).

BLOOD PIGMENT IN THE SALIVARY GLANDS

The salivary glands of *Rhodnius* (figure 1A) have been described by Baptist (1941); but this author curiously makes no reference to the fact that the principal gland is always distended with a bright cherry-red fluid. This fluid shows an absorption spectrum similar to that of oxyhaemoglobin but with the bands displaced towards the blue, being usually at about 570 and 532 $m\mu$. The α -band is the stronger, the β -band is slightly wider, and both are a little more diffuse than those of oxyhaemoglobin. Addition of ferricyanide causes no change. On adding hydrosulphite a single highly asymmetrical band appears; it extends from about 595 to 530 $m\mu$ with maximum intensity at about 575 $m\mu$. Sometimes a faint α -band of haemochromogen at 557 $m\mu$ is superimposed on this. On oxidation the pigment again shows the two bands as before. Carbon monoxide has no effect on the absorption spectrum.

This pigment does not seem so much altered as that circulating in the haemolymph. Its properties are intermediate between those of haemoglobin and parahaematin. It does not normally give haemochromogen on reduction with hydrosulphite. But it does so immediately if it has been stored for a day or so in glycerol; whereas laked blood still forms reduced haemoglobin even after storage in this way for several months. On treatment of the dried salivary secretion with Takayama's reagent, pink crystals of pyridine protohaemochromogen ($\alpha = 557 m\mu$, $\beta = 526 m\mu$) separate out. Of pigments previously described, that in the salivary glands seems to resemble haemalbumin most closely (Fairl y 1940).

In order to compare the concentration of pigments in the salivary glands with that in the blood of the mammalian host, about ten glands were placed on a waxed slide and dried with filter paper, due precautions being taken to prevent evaporation. They were then punctured with a needle and the contents rapidly taken up

into a waxed capillary pipette (Wigglesworth 1937), introduced into capillary tubing of 0.75 mm. bore and the ends sealed with paraffin. Blood from the rabbit was laked by freezing and thawing and introduced into an identical piece of capillary. Comparison showed that the salivary contents were markedly darker than the laked blood—containing perhaps 50 % more pigment.

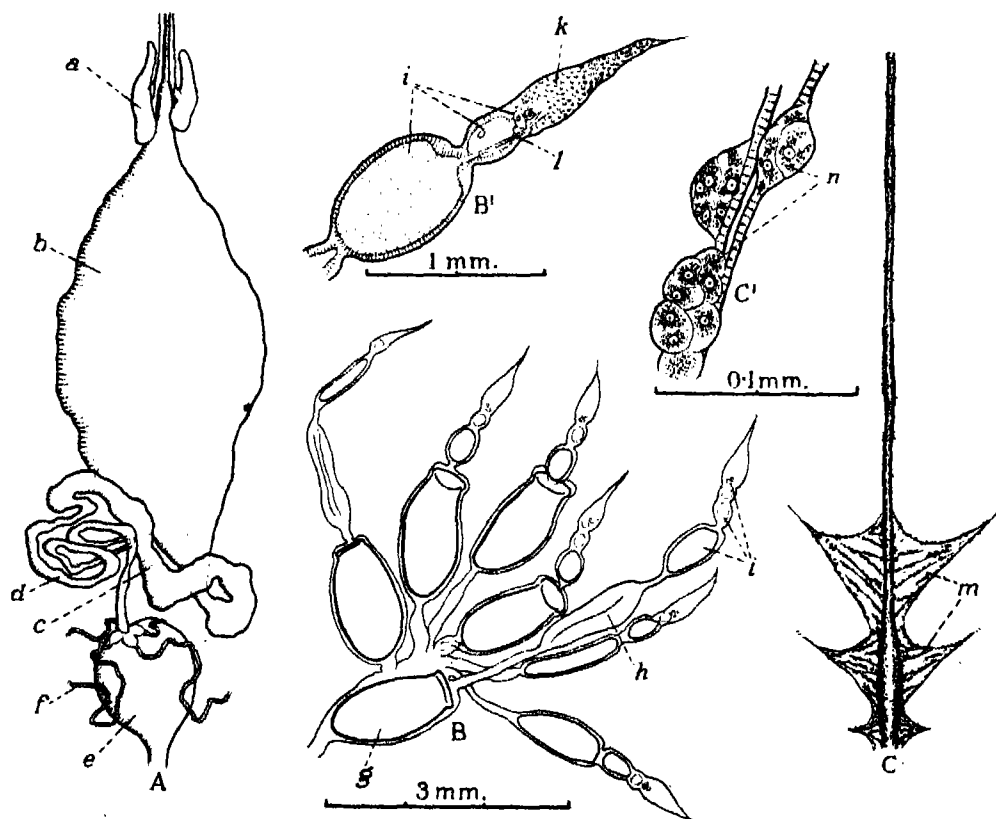


FIGURE 1. A, alimentary system of *Rhodnius*: a, salivary glands; b, stomach (mid-gut); c, anterior segment of intestine (mid-gut); d, posterior segment of intestine (mid-gut); e, rectum (hind-gut); f, Malpighian tubes. B, B', ovary: g, newly ovulated egg; h, corpus luteum; i, young oocytes; k, nurse cells; l, nutritive cord. C, C', heart and pericardial cells: m, alary muscles; n, detail of pericardial cells on the alary muscles, containing droplets of biliverdin.

BLOOD PIGMENT IN THE EGGS

As soon as the oocyte in the ovary of *Rhodnius* reaches a length of about 0.25 mm., and while it is still connected by the nutritive cord to the nurse cells (figure 1 B'), it acquires a pinkish tint. This grows much deeper as the oocyte, invested by the follicular cells, increases in size. If the fully developed egg, which measures rather more than 1.5 mm. in length, is immersed in glycerol and examined with the microspectroscope, it shows absorption bands like those of the haemolymph,

situated at about 567 and 530 $m\mu$. If the egg cap is removed and the pigment reduced with sodium hydrosulphite, it behaves like the pigment in the haemolymph and gives rise, at least in part, to haemochromogen (558 and 526 $m\mu$).

If newly laid eggs are crushed under a coverslip the yolk is seen to be full of refractile droplets of all sizes. But the pigment is diffused through the yolk plasma; it is not concentrated in any of the droplets. Its subsequent fate will be considered later (p. 328).

(As soon as the ectoderm is differentiated, its cells contain great quantities of an orange-red pigment which is quite unrelated to haemoglobin (Wigglesworth 1933.)

BILE PIGMENT IN THE PERICARDIAL CELLS

In the freshly dissected *Rhodnius* the blue-green colour of the pericardial cells throughout the dorsal vessel and along the alary muscles (figure 1c) is as striking a feature as the red salivary glands. This green pigment is present in the cells in the form of granules and droplets (figure 1c'). It is readily extracted with dilute alkali (0.1N sodium hydroxide). It shows diffuse absorption in the red but no definite absorption bands even in strong solutions. On adding minute traces of concentrated nitric acid (Gmelin's test) it becomes blue, then bluish violet, reddish violet, and finally reddish yellow. In the final stage it shows the absorption band of urobilin at 495 $m\mu$. On adding zinc acetate and iodine the green solution develops a slight red fluorescence, and the general absorption in the red becomes replaced by the intense narrow band of biliverdin at 532 $m\mu$ with a faint diffuse band at 585 $m\mu$ and a transient band at about 522 $m\mu$. The pigment is clearly *biliverdin*, possibly in the form of a complex with protein (chromoprotein). Its presence affords evidence that some at least of the haem in the haemoglobin that is absorbed is broken down.

DISTRIBUTION OF FREE IRON IN THE TISSUES

The distribution of free iron might be expected to give some indication of where such breakdown is taking place. This has been sought by fixation of the organs in Bouin's fixative and subsequent immersion in freshly prepared 1 % potassium ferrocyanide in 0.5 % hydrochloric acid. The same test has been applied to paraffin sections of the abdominal contents.

No deposits of iron can be detected in the normal insect in the ovaries, salivary glands, fat body, blood cells or Malpighian tubes. Small amounts may be present in the pericardial cells (p. 326). There are very large deposits in the wall of the gut. The stomach of the fully starved insect (which is normally distended with air) becomes coloured a distinct blue in the acid ferrocyanide. Microscopic examination in surface view shows that some of the blue colour is diffuse; but it is mainly concentrated in granules which are most abundant in cells scattered at intervals (figure 2A).

The heaviest deposits of iron are, however, in the anterior half or one-third of the intestine, where this forms a transverse loop across the abdomen (figure 1A). There may be some also in the coiled posterior half; but in many insects this segment is almost or completely free from iron. As in the stomach, the iron in the intestine tends to be concentrated in cells scattered at intervals. In the sections (figure 2B) it is seen to lie mainly between the nucleus and the apex of the cell. Some is diffused through the cytoplasm; but most is concentrated in spherical, oval or elongated deposits which colour intensely with ferrocyanide. If adjacent unstained sections, of which one only has been treated with ferrocyanide, are compared, it is seen that the iron-containing deposits appear, in their untreated state, as pale brown, greyish, or colourless inclusions, slightly refractile, spherical or irregular in outline.

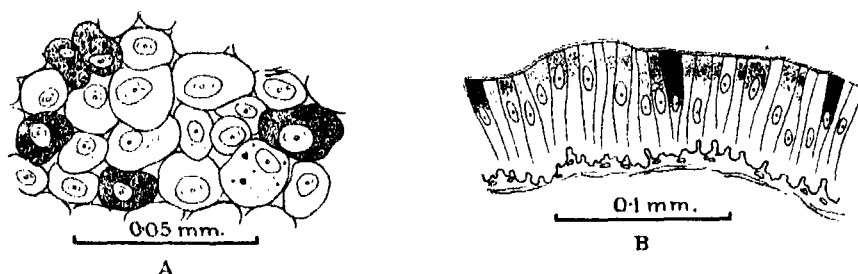


FIGURE 2. A, surface view of epithelium of stomach treated with potassium ferrocyanide showing free iron in the cells; B, section of intestine similarly treated.

BREAKDOWN OF HAEMOGLOBIN IN THE GUT EPITHELIUM

The distribution of free iron suggests that the breakdown of haemoglobin occurs largely in the epithelium of the gut. This is borne out by observations on the pigments in these cells. In the fully starved insect (say one month at 25° C. after a meal) the residue of the blood has disappeared completely from the lumen of the stomach and intestine. The stomach wall is green or brownish green; the anterior half of the intestine is usually green but may be slightly brownish in some insects; there is usually a middle segment that is deep brown, though this also may be green; the posterior part may show all shades from brown to green.

In sections, which are best mounted unstained or very faintly stained with orange G, these pigments are seen to have the same general distribution as the free iron (figure 3A). They also are limited to the apical region of the cells; indeed, they do not extend so far towards the base as do the iron deposits (figure 3A'). They are in the form of minute spherical granules or droplets which show all gradations from deep brown, through pale brown, yellowish, greenish yellow and green, to a clear blue green. This last pigment is ultimately discharged through the striated border into the lumen where it may accumulate in droplets of irregular form and in considerable quantity (figure 3C). It often appears as though all the green pigment

in a given cell were discharged about the same time; for where the newly excreted pigment is seen still adhering to the striated border (as in figure 3B) the cell beneath is commonly devoid of pigment.

The blue green pigment is clearly biliverdin. Extracts of the gut with dilute alkali at this stage show a brilliant Gmelin reaction with the formation of urobilin which can be demonstrated spectroscopically. This process of biliverdin formation and discharge in the gut cells of *Rhodnius* is very like that described and figured by Diwany (1919) in the chorionic epithelium of mammals, particularly of the dog.

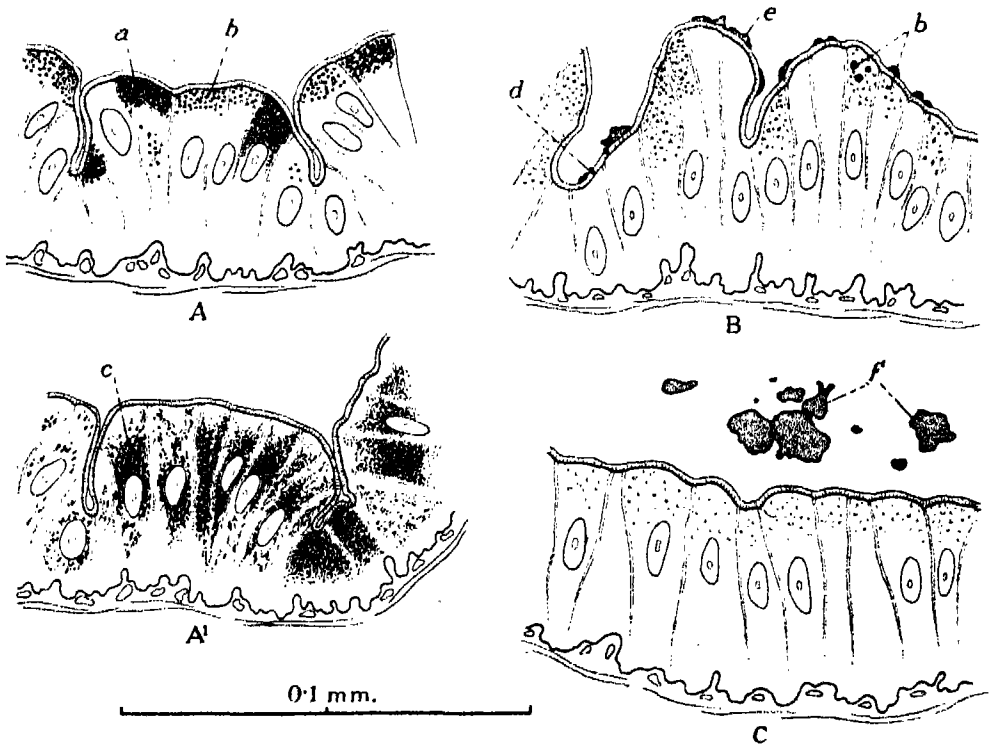


FIGURE 3. A, section of intestine of *Rhodnius*, unstained; A', adjacent section treated with potassium ferrocyanide; B, section of intestine showing discharge of biliverdin from the cells; C, section showing much biliverdin in the lumen. a, brown pigment; b, green pigment in the cells; c, iron deposits; d, droplet of biliverdin passing through the striated border; e, biliverdin adherent to surface of epithelium; f, biliverdin deposits in the lumen.

INTERMEDIATE STAGES IN HAEMOGLOBIN BREAKDOWN

We have seen that the dark brown residue in the lumen of the intestine is free protohaematin (p. 315) which on treatment with Takayama's reagent readily yields crystals of pyridine protohaemochromogen. If this protohaemochromogen is exposed to the air and warmed in Takayama's reagent, it gradually becomes deep green and shows the absorption spectrum of pyridine verdohaemochromogen: an

intense narrow band in the red at about $650\text{ m}\mu$ with fainter bands in the green at 535 and $500\text{ m}\mu$.

Now according to Lemberg (1937) verdohaemochromogen is to be regarded as one of the chief intermediates between haemoglobin and biliverdin, and it was expected that its formation might be demonstrated in the gut wall of *Rhodnius*. But when the brown pigment in the gut cells is dried and treated with Takayama's reagent it never yields any pink crystals; and the pyridine haemochromogen that is obtained has the α -band at about $546\text{ m}\mu$ and the β -band at about $515\text{ m}\mu$ —definitely nearer the blue than the bands of pyridine protohaemochromogen (557 and $526\text{ m}\mu$). When this red solution is exposed to the air, even at room temperature, it soon becomes deep green; the bands of haemochromogen disappear, an intense narrow band appears in the red at about $640\text{ m}\mu$, there is a fainter band at about $528\text{ m}\mu$ and an indistinct band at about $490\text{ m}\mu$. This pigment clearly corresponds to verdohaemochromogen; but it is produced even more readily and seems more stable than that derived from protohaemochromogen, and the absorption bands are all displaced towards the blue.

It has not been possible, however, to demonstrate this pigment in the gut wall. If the bright green segments of the gut in the fully starved insect are extracted with 0.1 N sodium hydroxide or better with 0.25 N sodium carbonate in 50% glycerol, a green solution is obtained which shows much general absorption in the red due to biliverdin but no discrete band. On keeping such extracts in the capillary tube for 24 hr. an intense absorption band appears at about $632\text{ m}\mu$ with a faint diffuse band at about $582\text{ m}\mu$ but with no bands in the green. This absorption spectrum remains unchanged on the addition of pyridine water, sodium hydrosulphite, or Takayama's reagent. It is immediately destroyed by sulphuric acid and no new bands develop.

This pigment is probably an oxidation product of biliverdin (biliviolin) arising after extraction; for its absorption spectrum, which agrees very closely with that of the zinc biliviolin complex (p. 318) cannot be seen in the fresh tissue before extraction but develops and increases in intensity in the extract over a period of some days.

On the other hand, if the green segment of the intestine is slit open a few days after feeding, the contents washed away with Ringer's solution and the gut wall ground up in 0.25 N sodium carbonate in 50% glycerol, it is possible to see a rather indefinite absorption band around $675\text{ m}\mu$, not easily separable from the general absorption in the red. On adding sodium hydrosulphite it is replaced by a distinct band at about $630\text{ m}\mu$.

Thus it has not been possible to demonstrate in the gut wall either verdohaemochromogen derived from normal haem or the corresponding pigment derived from the altered haem which enters into the composition of the brown pigment in the gut epithelium. But the green pigment present in the gut cells of the recently fed insect resembles rather closely the choleglobin of Lemberg, Legge & Lockwood (1939, 1941), another bile pigment precursor produced by the coupled aerobic

oxidation of haemoglobin and ascorbic acid. Reduced choleglobin has an absorption band at about $630\text{ m}\mu$; in the oxidized state the band shifts to $670\text{ m}\mu$. The pigment in *Rhodnius* does not seem, however, to be identical with choleglobin; it is certainly distinct from the 'pseudo-haemoglobin' of Barkan & Schales (1938) and the 'verdoglobin' of Havemann (1941).

It has not been possible to determine the true nature of the brown pigment. As we have seen, it yields a pyridine haemochromogen with bands at about 546 and $515\text{ m}\mu$. Treated in the same way as the haematin in the excreta (p. 315), it yields an acid porphyrin with bands at 593 , 572 and $548\text{ m}\mu$, and an alkaline porphyrin with bands at 617 , 568 , 540 and $503\text{ m}\mu$. Of the porphyrins commonly met with in nature these values agree most closely with those of coproporphyrin (Fischer 1926). But it is not possible to identify porphyrins with any certainty by spectroscopic means alone.

BREAKDOWN OF HAEMOGLOBIN IN THE PERICARDIAL CELLS

The question arises whether some breakdown of haemoglobin takes place also in the pericardial cells, or whether the biliverdin which these contain has been set free from the gut epithelium into the haemolymph and thence collected by the pericardial cells and stored.

Evidence that an excess of haemoglobin introduced into the haemolymph can be broken down by the pericardial cells will be given later (p. 325); and since kathaemoglobin is normally present in the haemolymph, it seems highly probable that such breakdown is constantly taking place. But free iron is found as a rule only in very small amounts in the pericardial cells: it must be continually discharged to the haemolymph and taken up by the gut epithelium. And in young insects it has usually not been possible to detect haematin in the heart by spectroscopic examination in Takayama's reagent. In older insects, however, which have received three or six feeds of blood, and in which perhaps the breakdown of haemoglobin in the pericardial cells is less rapid, the cells do contain some brownish granules among the green, and examination of the heart immersed in Takayama's reagent then reveals a faint band at $546\text{ m}\mu$ —the α -band of the same pyridine haemochromogen as can be demonstrated in the gut wall.

If a number of hearts are extracted with 0.25 N sodium carbonate in 50% glycerol, the extract develops after some days a strong band at $632\text{ m}\mu$ and a faint band in the yellow at about $585\text{ m}\mu$. This is probably biliviolin arising in the extract by oxidation of the biliverdin. It develops more slowly than in extracts from the intestine (p. 321); perhaps the abundant iron in the latter extracts catalyses the oxidation process.

FATE OF IRON

Since no free iron can be detected in the excreta, the iron liberated from the degraded haem must presumably accumulate throughout life in the wall of the gut. A batch of newly moulted *Rhodnius* was divided into three lots. The gut from

one lot was fixed as soon as all the blood residue had been passed on to the rectum. The second lot was given three meals at intervals of a fortnight, starved until the gut was empty and then fixed. The third lot was given six meals at intervals of a fortnight and similarly treated.

On dissection the stomach and intestine were found to be somewhat grayer in the insects that had had multiple feeds; the bright green of the biliverdin was partially obscured by other pigmented inclusions. And treatment with potassium ferrocyanide, both in whole mounts and in sections, showed a progressive increase in the amount of free iron in the cells (figure 4). In the insects that had had six feeds the cells throughout the gut were filled with the dense blue deposit. There was no accumulation of iron in the pericardial cells nor in any other tissue apart from the stomach and intestine.

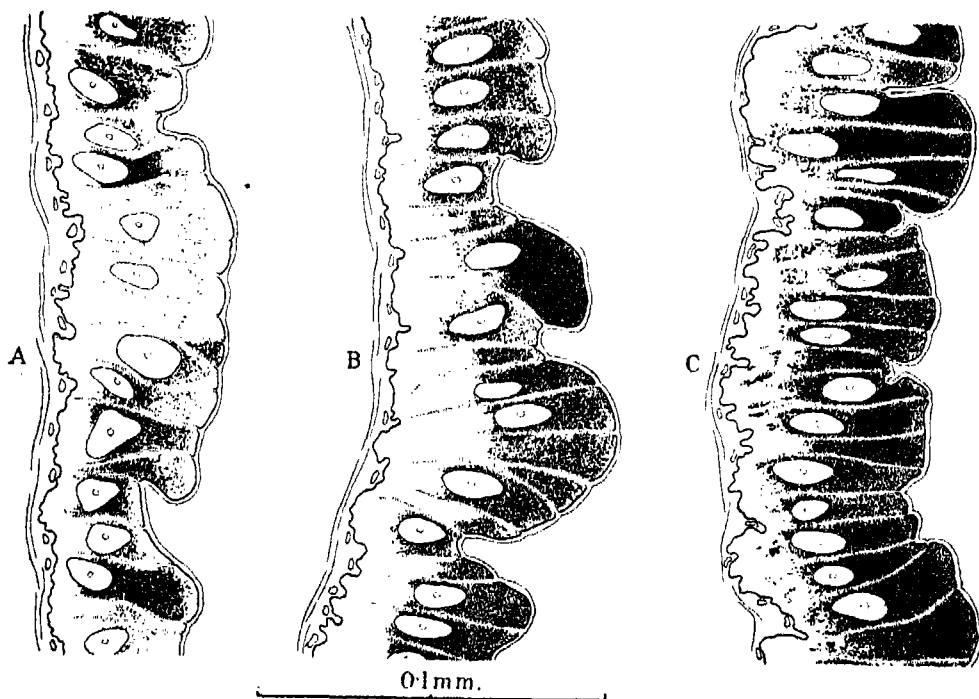


FIGURE 4. Longitudinal sections of the anterior segment of the intestine treated with potassium ferrocyanide. A, unfed adult; B, adult which had been given three feeds of blood; C, adult which had been given four feeds of blood and had received an injection of laked blood.

FATE OF BILIVERDIN

We have already seen that the biliverdin formed in the epithelium of the gut is ultimately discharged to the lumen (p. 319); and the experiments described in the preceding section have shown that there is no accumulation of biliverdin in the gut wall after multiple feeds of blood.

This same series of insects was also made use of in order to see whether there is a progressive accumulation of biliverdin in the pericardial cells. The dorsal abdominal segments, with the heart and alary muscles adhering to them, were removed and mounted whole after fixation in Carnoy's fixative (figure 1 c). There is a fair amount of variation in the quantity of biliverdin in the different individuals; but it was very obvious that the quantity increases after successive feeds: in the unfed adult the heart is pale blue-green; it is deep green after six feeds. It was evident also that the accumulation is greater in the males. This is doubtless due to the fact that much of the blood pigment entering the haemolymph in the female is taken up by the ovaries.

The next question is whether biliverdin suffers any further breakdown. No bilirubin has normally been detected in the pericardial cells. Granules of yellow pigment are sometimes seen in the gut epithelium, and among the deposits of biliverdin in the lumen of the gut there are elongated or irregular fragments of a brownish yellow pigment which appears to have been discharged from the cells. In the presence of the biliverdin it has not been possible to determine whether this pigment is bilirubin.

The urine of *Rhodnius* consists of uric acid spheres suspended in a deep yellow fluid (Wigglesworth 1931), and it seemed possible that this might contain bile pigment in solution. Urine free from faeces was collected from about twenty adult *Rhodnius* by blocking the anus with paraffin wax 24 hr. after feeding, keeping the insects for a week and then dissecting out the rectum, collecting the contents and filtering-off the uric acid. The filtrate gave no trace of Gmelin's reaction. The filtrate was saturated with ammonium chloride and refiltered. Then saturated with ammonium sulphate, acidified with sulphuric acid and shaken up with an equal volume of ether two parts and chloroform one part. No pigment was extracted by the ether-chloroform mixture and no absorption bands could be detected in the acidified aqueous solution. It must therefore be concluded that the urine contains neither bilirubin nor urobilin.

It is pointed out by Lemberg & Wyndham (1936) that in mammals bilirubin may be replaced by biliverdin if the glycogen in the liver becomes deficient, and that after injection of glucose biliverdin is reduced and bilirubin appears. It was interesting to repeat this experiment on *Rhodnius*. A number of adult bugs about four days after feeding were injected with about 10 mm.³ of saturated glucose solution. They were dissected two days later. The heart was now yellowish green and, on fixation and mounting, among the green deposits of biliverdin there were everywhere granules or elongated crystals of an orange pigment. It seems highly probable that this is bilirubin.

FATE OF HAEMOGLOBIN INJECTED INTO THE BODY CAVITY

From the foregoing observations it has been concluded that haemoglobin absorbed unchanged from the gut is denatured and oxidized to the ferric state

(kathaemoglobin) in the haemolymph and then taken up (i) by the ovaries without suffering further change, (ii) by the salivary glands, being modified again in the process, (iii) by the pericardial cells where it is converted to biliverdin, the iron being set free to the haemolymph, and (iv) by the gut epithelium where it goes by way of a brown pigment and a green pigment to biliverdin, the iron being stored. In order to test these conclusions the amount of haemoglobin circulating in the haemolymph has been artificially increased by injecting laked blood. About 10 mm.³, or in other experiments about 0.75 mm.³, of citrated blood laked by freezing and thawing were injected through the lateral pleat of the abdomen and the insects dissected at varying intervals.

Haemolymph. About 0.75 mm.³ of laked blood was injected two days after feeding. Six hours later the haemolymph, which was quite red, showed absorption bands at 577 and 540 m μ . On treatment with hydrosulphite there was a faint diffuse band of reduced haemoglobin and, superimposed on this, conspicuous haemochromogen bands at 557 and 526 m μ . Denaturation was already far advanced but not yet complete. Twenty-four hours after the injection the red haemolymph showed absorption bands at 573 and 536 m μ . After 4 days, their position varied in different individuals from 572 and 535 m μ to 568 and 532 m μ . They had now become very hazy and the β -band was at least as strong as the α -band. On adding hydrosulphite the haemochromogen bands alone were visible.

Thus it is clear that the injected haemoglobin is slowly denaturated and changed until it approximates to the pigment (kathaemoglobin) normally present in the haemolymph.

Salivary glands. In insects dissected 1 week after the injection of laked blood the contents of the salivary glands had a distinctly deeper red colour than those of controls from the same batch. This supports the belief that the pigment in the normal glands is derived from the blood pigment in the haemolymph.

Ovaries. As compared with the controls the entire ovary in the injected females, including nurse cells and follicular cells (figure 1 B, B'), shows a pink tint when dissected several days after the injection. The developing oocytes have a purplish or brownish tint instead of the usual pink. The egg after laying ranges from pale chocolate to a purplish brown.

Three weeks after the injection, when all the eggs have been laid, the ovaries still show a pinkish tint and the cellular plug which marks the position from which the 'corpus luteum' has been absorbed (Wigglesworth 1936) shows up as a reddish brown mass.

Pericardial cells. Meanwhile haemoglobin is slowly taken up by the pericardial cells. Twenty-four hours after the injection of 10 mm.³ of blood these are becoming enormously distended; the cytoplasm around the granules of biliverdin has a pink or brownish tinge and some brown granules are present. After two days the heart is becoming brown throughout. The cytoplasm of the pericardial cells is filled with brownish vacuoles and small granules of brown pigment. The biliverdin is becoming less. This process continues, the cells becoming filled with more and more brown

pigment in rounded deposits which grow progressively darker, and the biliverdin becoming less, until by the end of a week all the biliverdin has disappeared, the entire heart is deep brown and the distended cells are densely packed with small dark granules. At the end of a month there is no notable change.

In insects which had been kept alive for 3 or 4 months after the injection and given four feeds of blood during this period, many of the pericardial cells had disintegrated and disappeared. They were largely absent from the alary muscles and posterior region of the heart. Where they remained, some still contained only brown pigment but others showed only green pigment and in a few cells mixtures of brown and green could be seen.

The brown pigment which fills the pericardial cells one month after the injection appears to be identical with the brown pigment already described in the epithelium of the gut (p. 321). Extracted with dilute alkali, it shows a diffuse absorption band extending from about 620–590 $m\mu$. It yields a pyridine haemochromogen with bands at 546 and 515 $m\mu$, which will never crystallize out. This is readily oxidized to give a verdohaemochromogen with bands at 540, 526 and 490 $m\mu$. Treated with sulphuric acid the brown pigment yields a porphyrin with bands at 593, 572 and 549 $m\mu$. In alcoholic soda the bands are at 617, 508, 540 and 503 $m\mu$.

There was little evidence in these experiments that haemoglobin was being transformed into biliverdin in the pericardial cells. But the complete blockage of the cells by the pigment, the complete displacement of the existing biliverdin, and the subsequent disintegration of many of the cells, suggested that perhaps too much haemoglobin had been given. The experiment was therefore repeated, about 0.75 mm.³ of laked blood being injected in place of 10 mm.³

Under these conditions the displacement of biliverdin was never complete. In the male the heart became brown with a green tinge, but it remained green with a brown tinge in the female, in which so much of the blood pigment is being taken up by the ovaries (cf. p. 324). On treatment with potassium ferrocyanide the presence of abundant free iron could be demonstrated in certain of the pericardial cells. When the heart from such insects was mounted at intervals up to 3 months after the injection, the brown pigment could be seen to be becoming gradually less and the heart to be growing green again.

These results support the conclusion that the breakdown of haemoglobin follows the same lines in the pericardial cells as in the gut wall, and spectroscopic evidence of this is readily obtained. If four or five hearts are ground up with 0.25 N sodium carbonate at about 1 week after the injection, there is clearly visible, immediately after extraction, besides the absorption band of alkaline haematin at about 605 $m\mu$, a strong absorption band in the red at about 675 $m\mu$ and a faint band in the green at about 534 $m\mu$. On reduction with sodium hydrosulphite the band in the red shifts to about 630 $m\mu$. Here we have a pigment which is almost certainly of the verdohaem type and which, like that in the gut wall, approximates rather closely to the choleglobin of Lemberg *et al.* (1941).

Malpighian tubes. There are no obvious changes in the Malpighian tubes in the early days following injection. But at the end of 14 days traces of brown and green pigment can be seen in the cells of the upper segment.* After 3-4 weeks these cells usually contain numerous brown granules and here and there collections of greenish material. Biliverdin is now generally present in quantity in the lumen of the upper segment (figure 5A).

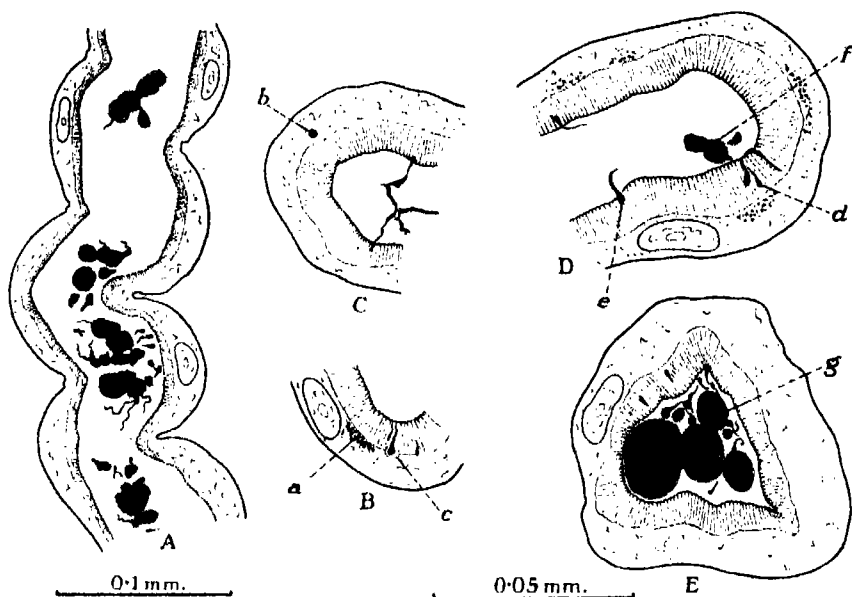


FIGURE 5. Excretion of biliverdin by Malpighian tubes. A, optical section of part of Malpighian tube showing biliverdin deposits in the lumen; B-E, transverse sections showing stages in the process of excretion. *a*, biliverdin granules accumulating in the cell; *b*, droplet of biliverdin in small vacuole; *c*, discharge of biliverdin into the striated border; *d*, droplet in form of filament passing through the border; *e*, discharge of filament to the lumen; *f*, fusion of such filaments to form droplets; *g*, masses of large droplets and filaments filling the lumen.

In sections and in whole mounts it is possible to follow the stages in the discharge of the biliverdin (figure 5B-E). It accumulates at the inner pole of the cell, among the numerous granules normally present in the upper segment (figure 5*a*). It then collects in a minute vacuole (figure 5*b*) which is discharged into the striated border (figure 5*c*) and the pigment passes as a filament, generally dilated towards its basal extremity (figure 5*d*) between the constituent elements of the border,† and so reaches the lumen (figure 5*e*). Here the filaments fuse to form irregular droplets (figure 5*f*) which may increase in size until they completely fill the tube (figure 5*g*).

* In the Malpighian tubes of *Rhodnius* the upper two-thirds secrete and the lower third reabsorbs (Wigglesworth 1931).

† The nature of the striated border has been discussed elsewhere (Wigglesworth 1931) and the conclusions reached confirmed by Zilch (1936).

In insects given four meals of blood after the injection and kept alive four or five months, the cells in the upper segments of the Malpighian tubes still contain many granules of brown pigment. Biliverdin may also persist in small quantities in the lumen.

It is probable that most of the biliverdin excreted by the Malpighian tubes in these experiments represents pigment that has been displaced from the pericardial cells as these become blocked with haematin. But the presence of brown and yellowish pigments in the cells, as well as the presence of small amounts of free iron, proves that there is some breakdown of haemoglobin in the cells of the Malpighian tubes themselves.

Alimentary canal. In insects starved for three to four weeks after the injection of haemoglobin the freshly dissected gut shows no very obvious differences from that of the normal insect. Brown, grey and green pigments are present as usual, but the bright green colour is generally obscured by a larger quantity of brownish grey. When treated with potassium ferrocyanide there is a most striking difference; the entire intestine shows an intense blue-black colour, the blue granules completely filling many of the cells (figure 4c). The stomach also is much darker than usual.

Muscles and other tissues. One month after the injection of haemoglobin many of the tissues, such as the fat body, tracheal epithelium, etc., have a slightly brownish tinge. Sometimes the fat body appears greenish. But there are no obvious deposits of haemoglobin derivatives, and free iron is nowhere present.

The thoracic muscles of *Rhodnius* are normally of a pinkish colour. The fresh muscle examined spectroscopically in a capillary tube after treatment with sodium hydrosulphite shows the bands of reduced cytochrome at about 604, 566, 550 and 520 $m\mu$. On the addition of pyridine the gap between the *b* and *c* bands at 566 $m\mu$ and 550 $m\mu$ is obscured by a strong protohaemochromogen band, showing the presence of free haematin or parahaematin in the muscle. If the muscle from an insect one month after injection of laked blood is compared with the normal, no difference in the intensity of either set of absorption bands can be detected.

FATE OF BLOOD PIGMENT TRANSMITTED TO THE EGG

If the first stage nymph of *Rhodnius* is dissected immediately after hatching, the yolk residue which fills the stomach still contains undigested blood pigment of a bright red colour, as though the insect had already had a feed of blood. The depth of colour, like that of the newly laid eggs, varies in different individuals. The intestine is empty or contains a trace of red pigment. The rectum is filled with clear colourless fluid with a few uric acid spheres (figure 6A).

On comparison with the contents of the salivary glands of the adult insect it can be seen that the pigment is more concentrated in the glands. It appears about equally concentrated in the stomach contents and in laked rabbit blood. Microscopically the stomach contents show small scattered fat droplets and numerous

large, apparently watery vacuoles free from pigment. The pigment is confined to the plasma between the vacuoles. Here it must attain a high concentration.

This concentrated pigment in the stomach of the newly hatched nymph resembles that occurring in dilute form in the haemolymph of the adult, and is probably kathaemoglobin. It shows somewhat hazy absorption bands at about 567 and 530 $m\mu$, the β -band being broader and darker. These bands are unaffected by potassium ferricyanide. On reduction with sodium hydrosulphite the intense absorption spectrum of protohaemochromogen (558 and 527 $m\mu$) immediately appears. On exposure to the air this slowly reverts to the spectrum of parahaematin. On adding Takayama's reagent to the dried substance, crystals of pyridine

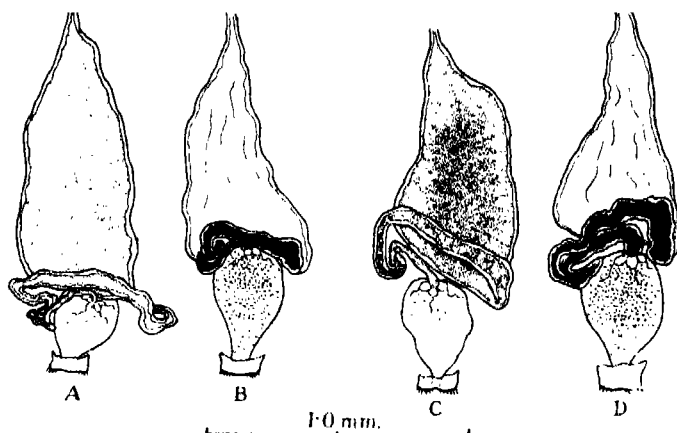


FIGURE 6. Alimentary canal of unfed first stage nymphs of *Rhodnius*. A, normal, immediately after hatching; stomach and intestine filled with brick red material. B, normal, 10 days after hatching; stomach empty, intestine containing residual haematin. C, from female injected with laked blood, immediately after hatching; stomach and intestine filled with deep red material. D, from female injected with laked blood, 10 days after hatching; stomach empty, intestine containing larger quantity of residual haematin.

protohaemochromogen separate out. The pigment differs from that in the salivary glands (p. 316) in its brick red colour, the greater haziness of its absorption bands with the β -band broader and more dense, and in its conversion into haemochromogen on reduction with hydrosulphite. The salivary glands of the newly hatched first stage nymph contain a clear colourless fluid without blood pigment.

During the next few days after hatching the pigment in the stomach is passed on to the intestine and digested. By ten days after hatching the stomach is empty; the intestine contains only the black residual haematin in the lumen, with traces of brown and green pigments in the epithelial cells; the rectum is filled with uric acid spheres in a deep yellow liquid (figure 6B). The salivary glands are now distended with red fluid. This pigment has the same properties as that in the salivary glands of the adult; it is not identical with the parahaematin present in the stomach on hatching. If these insects are allowed to feed, and dissected immediately afterwards, the salivary glands are found partially collapsed. Thus a

small part of the blood pigment removed by the parent from the host is ultimately reinjected by the offspring.

In the first stage nymphs that emerge from the brownish eggs laid by females which have been injected with laked blood,* the stomach contains a still more concentrated solution of blood pigment (figure 6c); and at the end of ten days, when this has been completely digested, the intestine contains notably more haematin (figure 6d).

FATE OF HAEMOGLOBIN IN SOME OTHER BLOOD-SUCKING ARTHROPODS

Preliminary observations have been made on a number of other blood-sucking insects and ticks. These show great differences even among closely related forms.

INSECTA

HEMIPTERA

Triatoma infestans. This is a blood-sucking reduviid related to *Rhodnius*. Digestion in the intestine follows the same lines as in *Rhodnius*. The *haemolymph* is colourless and no haemoglobin derivatives can be detected in it. The *salivary glands* are distended with clear colourless fluid. The *eggs* are very pale pink, almost white when newly laid, and by the tests employed no blood pigment can be found in them; nor is there any haematin in the gut of the unfed first stage nymph. The *pericardial cells* are of a very pale reddish brown colour. They appear to contain parahaematin. Examined fresh in a capillary tube they show very faint and hazy bands at about 570 and 530 $m\mu$. On adding sodium hydrosulphite the sharp bands of protohaemochromogen appear at 558 and 527 $m\mu$. The *stomach* and the anterior part of the *intestine* show no green or brown pigments in the cells; at the hind end of the intestine the epithelium does contain brown pigment which, on treatment with Takayama's reagent, like the haematin in the lumen of the gut, gives rise to pyridine protohaemochromogen. The *Malpighian tubes* contain a fair amount of bright green pigment in the cells of the upper segment. This is Gmelin positive and gives rise to urobilin. The uric acid in the lower segments has much yellow pigment mixed with it so that it appears bright yellow. This pigment is Gmelin negative. On treatment with acid ferrocyanide the Malpighian tubes show a few scanty blue granules among the green. No deposits of free iron can be seen in any other tissues; the stomach and intestine, in striking contrast with *Rhodnius*, remain quite colourless.

After the injection of laked blood a large amount of brown pigment appears in the pericardial cells. This yields only protohaemochromogen and not the altered haemochromogen formed in *Rhodnius*.

* There seems to be a far higher mortality in these eggs than in normal eggs from the same insects. Perhaps the blood pigment, which cannot be used for tissue building in the developing egg, has displaced too much of the normal protein of the yolk.

Triatoma brasiliensis. Here again the *haemolymph* is colourless and contains no demonstrable haemoglobin derivative. The *salivary glands* are colourless. The *pericardial cells* are brilliant blue green, containing even more pigment than in *Rhodnius*; this pigment gives an intense Gmelin's reaction. When the cells are treated with ferrocyanide, blue-black granules appear among the green. The *eggs* are very pale pink when newly laid and no blood pigment can be detected in them; nor can haematin be detected in the intestine of the young nymph after hatching. The epithelium of the *stomach* and the anterior half of the *intestine* are quite colourless. In the cells of the coiled posterior half of the intestine there are brown and green pigments mixed in varying degrees. The brown pigment on treatment with Takayama's reagent yields a haemochromogen with bands at 550 and 520 $m\mu$. The *Malpighian tubes* show traces of green pigment. There are minute quantities of free iron in them; rather larger deposits in the wall of the hind end of the intestine, but none in the stomach or anterior half of the intestine.

A number of *T. brasiliensis* were injected with laked blood. Six days later the haemolymph showed absorption bands at 577 and 540 $m\mu$. On reduction with hydrosulphite it showed bands of protohaemochromogen at 558 and 527 $m\mu$. The globin had become denatured, but there was not the shift in the bands associated with parahaematin formation as in *Rhodnius*.

The pericardial cells of these injected insects contained a mixture of green and brown pigments. When treated with Takayama's reagent, the resulting haemochromogen shows intense bands with centres at 550 and 520 $m\mu$ like that from the intestine. Whether this brown pigment is a modified haematin different from that in *Rhodnius* or whether it is a mixture of the same haematin as in *Rhodnius* with protohaematin is uncertain. It yields an acid porphyrin which likewise has bands in an intermediate position (594, 574 and 550 $m\mu$).

Green pigments of more than one type have been demonstrated spectroscopically in the pericardial cells of *Triatoma brasiliensis*. The normal heart extracted with 0.25 N sodium carbonate develops a single narrow band with centre at 640 $m\mu$, unaffected by Takayama's reagent. After injection of laked blood and extraction with dilute sodium carbonate two types of absorption spectrum have been found: (i) There may be a single band at about 630 $m\mu$ which is unaffected by Takayama's reagent. (ii) There may be two very distinct bands in the red, one at about 675 $m\mu$, the other, due to alkaline haematin, at about 605 $m\mu$, and a faint band in the green at about 534 $m\mu$. On reduction with hydrosulphite the band in the red shifts to about 630 $m\mu$. This is clearly the same pigment, resembling choleglobin, that was demonstrated in the pericardial cells of *Rhodnius* (p. 326).

Eutriatoma sordida. Very few specimens of this species have been available for study. The *haemolymph* is colourless and no haemochromogen can be detected in it. The *salivary glands* are colourless. The *pericardial cells* are of a yellowish green colour. After soaking in glycerol followed by treatment with fuming nitric acid they at once become bright blue green throughout, later passing through the usual sequence of colours until the yellow pigment showing the urobilin band is produced.

Microscopic examination of the pericardial cells shows that many of them contain orange granules among the green. On treatment with nitric acid these also become blue green. It is probable therefore that the nephrocytes contain bilirubin as well as biliverdin. The cells of the upper segments of the *Malpighian tubes* are filled with bright blue green pigment, Gmelin positive. In the lower segments the lumen is filled with dead white uric acid. No green pigment is present in the cells of the *stomach* or *intestine*. The eggs are very pale pink when newly laid. They do not appear to contain any blood pigment.

Cimex lectularius. The *salivary glands* of the bed-bug, described by Patton & Cragg (1913) as 'slightly yellowish', are brownish red in colour when freshly dissected. This pigment shows absorption bands at 578 and 542 $m\mu$. The α -band is no darker than the β -band and the β -band is much broader. On addition of hydrosulphite a single diffuse band with centre about 575 $m\mu$ is produced. This pigment, like that in the salivary glands of *Rhodnius*, resembles haemalbumin (Fairley 1940).

No pigment can be seen in the *pericardial cells*. Digestion in the lumen of the gut proceeds as in *Rhodnius* but no brown or green pigments can be seen in the epithelium of the *stomach* or *intestine*. The *Malpighian tubes* are colourless, and no deposits of free iron can be demonstrated in any of the tissues.

The *eggs* are of a pale buff colour and as development proceeds and the yolk residue becomes concentrated in the lumen of the stomach a very faint diffuse band of alkaline haematin at about 610 $m\mu$ and a very faint band at about 550 $m\mu$ are detectable in this part of the living egg. On removal of the egg cap and immersion in sodium hydrosulphite there is no change; in Takayama's reagent the spectrum of protohaemochromogen becomes very obvious. In the *newly hatched nymph* the contents of the stomach are brownish red and show the spectrum of alkaline haematin; those of the intestine are dark brown. After a few days there remains only a small black residue in the intestine. Immediately after hatching, the salivary glands, which can be readily observed in the living bug, are only faintly coloured. As digestion in the gut proceeds they darken (as in *Rhodnius*) and their reddish contents show the same properties as the pigment in the glands of the adult *Cimex*.

After the injection of laked blood the pericardial cells become filled with large droplets of brown pigment. Treated with Takayama's reagent these show the absorption bands of pyridine protohaemochromogen. No green pigment can be seen. The cells of the upper segments of the Malpighian tubes also contain brown pigment. The eggs laid by the injected females are darker in colour than the normal eggs. If the cap is removed and the egg contents pressed out into sodium hydrosulphite no haemochromogen bands are formed. Intense bands of protohaemochromogen appear on adding Takayama's reagent. The quantity of haematin present in the intestine of the nymphs hatching from these eggs is obviously greater than in normal controls.

ANOPLURA

Pediculus humanus corporis. In the body louse, which feeds every few hours, the blood goes straight to the stomach where it is digested. Large quantities are passed on to the hind gut and discharged before digestion is complete. The excreta taken up in distilled water therefore show a strong methaemoglobin band at $630\text{ m}\mu$ with bands of oxyhaemoglobin at 577 and $540\text{ m}\mu$. On treatment with sodium hydro-sulphite reduced haemoglobin is formed. No free iron is usually present in the excreta.

The *haemolymph* is colourless. The tubular *salivary glands* are colourless; the reniform pair are slightly pinkish, but no absorption bands can be detected either in the fresh glands or after immersion in Takayama's reagent. The heart is colourless but, as observed by Nuttall & Keilin (1921) the peri-oesophageal and disseminated nephrocytes, which correspond with the *pericardial cells*, contain droplets of greenish pigment. Treated with nitric acid after clearing in glycerol this pigment becomes violet and then fades to pink and yellow; it is probably biliverdin. There is no brown or green pigment in the *epithelial cells of the gut*, and the *Malpighian tubes* are colourless. Treatment with acid ferrocyanide reveals no large deposits of *free iron* in any of the tissues; there are, however, some fine blue granules among the green inclusions in the nephrocytes. The developing *eggs* in the ovary are pink. On puncturing the egg and immersing in Takayama's reagent a distinct band of protohaemochromogen becomes visible at $557\text{ m}\mu$.

On hatching from the egg the stomach of the young louse contains a brownish red residue. If it is starved during the next 24 or 48 hours the contents become green. They then give a brilliant Gmelin's reaction, in the later stages of which the absorption band of urobilin is easily detectable. This biliverdin, and the brown pigment, are limited to the lumen of the gut. If the gut, containing a mixture of green and brown pigment, is treated with Takayama's reagent, crystals of pyridine protohaemochromogen (557 and $526\text{ m}\mu$) are at once formed. There is no trace of the absorption band at $546\text{ m}\mu$ given by the altered haematin which appears to be an intermediate stage in the formation of bile pigment in *Rhodnius*. If the gut of the starved, recently hatched nymph is treated with potassium ferrocyanide, it is seen that the *free iron* is confined to the lumen of the stomach and hind-gut; there are no deposits in the cells.

Stomachs were dissected out from 150 recently hatched nymphs and placed in 0.5 mm.^3 of 0.5 N sodium carbonate in 50% glycerol. A faint band in the red at about $530\text{ m}\mu$ became visible, as well as the band of alkaline haematin at about $610\text{ m}\mu$. As in *Rhodnius* (p. 321) the band in the red may have been due to a verdohaem pigment or to a biliviolin formed in the extract.

In the adult louse the partially digested blood is normally passed on and excreted before any biliverdin is formed; but if the insect is starved for a day or two it is possible to see some bright green, Gmelin-positive granules among the brown contents of the mid-gut.

Several days after the injection of *Pediculus* with laked blood, brown droplets become plentiful in the peri-oesophageal and disseminated nephrocytes. In Takayama's reagent these yield protohaemochromogen. Ten days after injection there is probably an increased quantity of biliverdin in these cells. (There is, however, much individual variation in the normal insects.) The Malpighian tubes remain colourless. The eggs are of a much deeper red than usual. Examined in the ovary they show distinct bands at 577 and 540 $m\mu$. On treatment with hydrosulphite the bands fade but no haemochromogen is formed until alkali is added. Many of the most deeply coloured eggs fail to hatch; when hatching does take place there is a notable increase in the brown and green contents of the stomach of the young larva.

SIPHONAPTERA

Nosopsyllus (Ceratophyllus) fasciatus. In the rat flea the blood goes directly to the stomach where it is digested. As in the louse, digestion is very incomplete, and the faeces contain much unchanged haemoglobin. No bile pigment formation can be detected in the gut. The epithelial cells of the stomach and Malpighian tubes are colourless. The salivary glands are almost colourless and no haemochromogen can be demonstrated in them. The same applies to the eggs and the pericardial cells.

DIPTERA

Anopheles maculipennis and *Aedes aegypti*. In these mosquitoes digestion in the gut seems to be very complete and only a small black residue of protohaematin remains in the faeces. No formation of bile pigment can be detected. The salivary glands, pericardial cells and Malpighian tubes are colourless. The yolk of the eggs is creamy white and no haemochromogen can be detected in it.

ARACHNIDA

Ornithodoros moubata. Digestion in the tick *Ornithodoros savignyi* has been described by Christophers (1906). The blood is slowly digested in the capacious stomach with its diverticula. Haemoglobin is taken up by swollen epithelial cells of the gut wall which subsequently become detached and set free into the lumen. In these cells the blood pigment is converted into intensely black globules 0.5–5 μ in diameter which are ultimately discharged into the cavity of the gut. In addition, minute dark granules are dispersed throughout the small cells of the gut wall. No black pigment can be seen in the other tissues or in the body cavity.

In the egg laying female of *O. moubata* the *haemolymph* has a very pale buff colour; no absorption bands are detectable. On adding Takayama's reagent a very faint band can be seen at 557 $m\mu$. The *gut contents* are red brown with a black deposit of haematin. Taken up in Ringer's solution they show only the intense absorption bands of oxyhaemoglobin. On treatment with Takayama's reagent, the large granules in the distended cells of the *gut wall* and the fine dark granules

in the smaller cells immediately give crystals of pyridine protohaemochromogen. No free iron can be detected in the gut lumen or cells or in any other tissues. No nephrocytes containing haemoglobin derivatives can be found.

The eggs of *Ornithodoros* are reddish brown in colour. A single egg examined spectroscopically shows a faint diffuse band of alkaline haematin at about 612 $m\mu$. On treatment with Takayama's reagent this is converted into protohaemochromogen. If the newly hatched larva is extracted from the remains of the egg shell, the sacculated gut is seen to be filled with a reddish brown fluid. The larva of *O. moubata* moults to the nymph without feeding, and this will resist starvation for months. At the end of one month many of the cells of the gut of the unfed nymph contain quantities of deep brown or black granules.

Ornithodoros delanoei acinus. The eggs of this large tick are a much deeper brown than those of *O. moubata*. A single egg shows an intense absorption band of alkaline haematin.

Ixodes ricinus. Observations have been made on the engorged egg-laying female of the sheep tick. The haemolymph is pinkish brown in colour and shows absorption bands at 620, 538 and 500 $m\mu$. On reduction with sodium hydrosulphite the haemolymph becomes red, the band at 620 $m\mu$ fades and new bands immediately appear at 576 and 540 $m\mu$. These resemble the bands of oxyhaemoglobin though they are rather hazy and they remain unchanged in the presence of a large excess of hydrosulphite. On adding sodium hydroxide strong bands of protohaemochromogen (558 and 527 $m\mu$) appear. This altered pigment of the haemolymph recalls the pseudomethaemoglobin or methaemalbumin formed when haemoglobin is incubated with human or simian plasma (Fairley 1940).

Digestion in *Ixodes* has been described by Nordenskiöld (1908) and Samson (1909). The contents of the gut and diverticula consist of a core of unaltered blood with a covering of granular black haematin. In distilled water they show only the sharp bands of oxyhaemoglobin. The cells of the gut wall contain numerous black granules of haematin yielding protohaemochromogen. No green or yellow pigments can be seen and there are no deposits of free iron.

Nephrocytes containing brownish pigment were described by Bruntz (1910) lying ventral in the body cavity. These are filled with large droplets ranging in tint from yellow and brown to almost black. Gmelin's test was negative. No free iron could be demonstrated in them* and on treatment with Takayama's reagent the deposits of all shades become pink and show the spectrum of pyridine protohaemochromogen. Free iron was absent also from the other tissues.

The eggs in the ovary are pale reddish brown. If a capillary tube is packed with them the band of alkaline haematin with its centre about 610 $m\mu$ is easily seen. The newly hatched larva has only pale red brown contents in the lumen of the gut. After starving for nine months the sacculated gut is shrunken and cells and lumen contain much black pigment.

* In preliminary observations Diwany (1919) states that the nephrocytes of *Ixodes ricinus* are filled with deposits of free iron.

DISCUSSION

Haemoglobin breakdown in different species. In all the blood-sucking arthropods studied the great bulk of the haemoglobin is broken down in the lumen of the gut to haematin, which is then excreted unchanged. But there are striking differences, even in closely related species, in the fate of the relatively small quantities of blood pigment that are absorbed without being digested.

At the one extreme are the mosquitoes and the flea, in which there is no evidence of any absorption of blood pigment. All the other forms studied show absorption in varying degrees. Thus pigment circulates in the haemolymph of *Rhodnius* in the form of kathaemoglobin (parahaematin); in *Ornithodoros* probably as alkaline haematin; in *Ixodes* in the form of a pigment resembling methaemalbumin.

In the bed-bug *Cimex* and in the ticks *Ixodes* and *Ornithodoros*, the fraction absorbed is broken down no further than protohaematin. In the louse *Pediculus* and the reduviids *Rhodnius*, *Triatoma brasiliensis* and *T. infestans* it is degraded to biliverdin, and in *Eutriatoma sordida* probably to a mixture of biliverdin and bilirubin. In no species were free prophyryns observed.

The formation of the pigments takes place in the nephrocytes (pericardial cells) and in the lumen of the gut in *Pediculus*, in the Malpighian tubes in *Triatoma infestans*, in the pericardial cells, Malpighian tubes and the hind end of the intestine in *T. brasiliensis* and *Eutriatoma sordida*, and in the pericardial cells, throughout the whole length of the stomach and intestine, and in the Malpighian tubes in *Rhodnius*. This process takes place on the largest scale in *Rhodnius* and only in this insect are there massive accumulations of free iron in the epithelium of the gut wall, and the excretion of large amounts of biliverdin into the intestine.

Combined protohaematin is transmitted to the egg in many species: as unchanged oxyhaemoglobin in *Pediculus*; as alkaline haematin in *Cimex* and in the ticks *Ornithodoros* and *Ixodes*; as parahaematin in *Rhodnius*. There seems to be no such transfer in *Triatoma*.

Combined protohaematin collects in the salivary glands of *Rhodnius* and *Cimex* in the form of a pigment resembling haemalbumin (Fairley 1940). There is no such transfer in the other insects.

Breakdown of haem. From the standpoint of comparative biochemistry particular interest attaches to those insects in which bile pigments are produced. In recent years haemoglobin has been converted into biliverdin in the laboratory by a number of different procedures. As an intermediate stage in this process dark green pigments are produced in which the iron-porphyrin ring of the prosthetic group is believed to have been opened but to be still held together by the iron atom, now labile and readily detached by acids. The protein or the nitrogenous substance united to this group varies, and the prosthetic group itself may not be constant. Such pigments are verdohaemochromogen (Lemberg 1937), choleglobin (Lemberg *et al.* 1941), pseudohaemoglobin (Barkan & Schales 1938), verdoglobin (Havemann 1941). Prosthetic groups of this open ring, bile pigment type are believed

to occur naturally in catalase (Lemberg & Wyndham 1936) and cytochrome (Lemberg 1937). In the formation of bile pigments by the tissues it is believed that more than one mechanism may be concerned (Drabkin 1942).

The green pigments with absorption bands in the red which have been demonstrated in extracts of the pericardial cells and intestinal epithelium of *Rhodnius* (p. 321) and in the pericardial cells of *Triatoma brasiliensis* (p. 331) are of several kinds. Some are probably oxidation products of biliverdin (biliviolins); but others are almost certainly verdohaem compounds not identical with any previously described but closely resembling choleglobin.

It is of interest that in the two insects in which the formation of bile pigments is most obvious (*Rhodnius* and *Triatoma brasiliensis*) the haematin in the cells alongside the green pigments is no longer protohaematin. This suggests that a change in the side chains of the iron porphyrin may precede the opening of the ring. On the other hand, no such altered haematin can be found in the newly hatched *Pediculus* when biliverdin is being produced in the lumen of the gut. And the persistence of this brown pigment, apparently indefinitely, in the gut wall of starved insects and in the pericardial cells and Malpighian tubes after the injection of laked blood, suggests that it may perhaps be just the end product of some side reaction.

The present paper is intended only as a rough outline of this aspect of insect physiology. It is obvious that the whole problem of haemoglobin breakdown in this material would well repay detailed study by modern methods.

Function of pericardial cells. The pericardial cells of insects were long regarded as storage kidneys (Kowalevsky 1892; Bruntz 1910). In recent years, because of their capacity for collecting from the blood and retaining in their cell bodies colloidal particles of dyes, proteins, etc. they have been compared with the reticulo-endothelial system of vertebrates (Poll 1934). In the breakdown of haemoglobin to bile pigments in vertebrates, free iron and biliverdin both appear first in the Kupffer cells of the liver (McNec 1913) and the evidence points to reticulo-endothelial cells in some part of the body being always responsible for their formation (Rich 1925). The parallel between the pericardial cells in *Rhodnius*, for example, and the reticulo-endothelial cells of vertebrates is therefore very close.

Role of gut wall in metabolism. Recent work on the mosquito larva (Wigglesworth 1942) has shown that the gut wall plays an important part in intermediary metabolism. It has, for example, a glycogenic function. The present work shows that in *Rhodnius* (though not in the other insects studied) it plays an even greater part than the pericardial cells in the breakdown of haemoglobin to biliverdin and that it is the main seat for storage throughout life of the liberated iron. In these respects, as in the intermediary metabolism of the mosquito larva, the same functions can be performed by the cells of the Malpighian tubes.

Waterhouse (1940) has described in the middle segment of the gut of blow-fly larvae a small band of cells containing masses of free iron. This was regarded as iron absorbed from the gut. But it was pointed out by Lennox (1940) that it might conceivably be derived in part from degraded haem.

Transfer of blood pigment to the egg and salivary glands. It is usually supposed that the constituents of the egg yolk are formed by the nurse cells and follicular cells. But the fact that haemoglobin (apparently unchanged in *Pediculus*, as kathaemoglobin in *Rhodnius*) is transferred to the oocyte, suggests that the normal protein constituents may perhaps be formed elsewhere and merely transmitted to the yolk by these cells. In *Rhodnius* at least, the globin has become denatured during circulation in the haemolymph and in the process of transfer to the egg. But the failure of the protein to undergo any further breakdown until the gut of the young insect becomes functional and digests the yolk residue, indicates how limited are the powers of the tissues, even during embryonic development, for dealing with foreign proteins once they have passed the epithelium of the gut. That haemoglobin should pass through the cells in the way described need cause no surprise. It is well recognised in mammals that there is a constant flow of protein to and fro between the tissues and the blood and that haemoglobin is constantly being filtered through the kidney glomerulus and reabsorbed by the renal tubules (Yuile 1942).

It is almost certain that the pigment in the salivary glands in *Rhodnius* is derived from that in the haemolymph. The prosthetic group is unchanged; but the properties of the pigment are quite different (p. 316): the protein has clearly been altered and is no longer denatured.

The extent to which haemoglobin is absorbed unchanged varies greatly in closely related species. Only in *Rhodnius* and *Cimex* does transfer to the salivary glands take place. These facts recall the specificity in the transmission of viruses by plant-sucking insects in which the determining factor in some instances (operative sometimes between mutant strains of the same species) is the ability of the virus to pass through the gut wall of the insect host (Storey 1933).

Apart from the haemoglobins (erythrocrucorins) that are functional respiratory pigments in some chironomid larvae, haemoglobin has a curiously sporadic distribution in insects. It occurs in the so-called 'tracheal cells' of *Gastrophilus* (Radu 1932) and certain aquatic Hemiptera (*Buenoa* (Hungerford 1922) and *Anisops* (Poisson 1926)) and in the accessory genital glands of the male in the aquatic bug *Macrocorixa* (Brindley 1929). No function for these pigments has been demonstrated. Perhaps, like the pigment in the salivary glands of *Rhodnius* and *Cimex*, they are functionless inclusions derived from the blood of the host or prey of the insects that contain them. This explanation is least probable in *Macrocorixa* which is mainly, though not solely, herbivorous.

The *Ornithodoros delanoei* were kindly supplied by Mr R. J. Whittick of the British Museum (Natural History). It is a pleasure to acknowledge also the large amount of preparatory work done by Miss W. Wall.

REFERENCES

- Baptist, B. A. 1941 *Quart. J. Micr. Sci.* **82**, 91.
- Barkan, G. & Schales, O. 1938 *Hoppe-Seyl. Z.* **253**, 83.
- Brindley, M. D. H. 1929 *Trans. Ent. Soc. Lond.* p. 5.
- Bruntz, L. 1910 *Ann. Sci. Nat. (Zool.)*, **12**, 265.
- Christophers, S. R. 1906 *Sci. Mem. Med. Sanit. Dep. India* (n.s.), no. 23.
- Diwany, H. F. 1919 *Étude histologique de l'embryotropie hématique des mammifères et du tube digestif de quelques invertébrés hématophages*. Paris.
- Drabkin, D. L. 1942 *Ann. Rev. Biochem.* **11**, 531.
- Fairley, N. H. 1940 *Brit. J. Exp. Path.* **21**, 231.
- Fischer, H. 1926 *Handb. biol. ArbMeth. Abt. Chem. meth.* **11**, 169.
- Fischer, H. & Treibs, A. 1926 *Tabul. biol., Berl.*, **3**, 339.
- Havemann, R. 1941 *Biochem. Z.* **308**, 1.
- Hungerford, H. B. 1922 *Canad. Ent.* **54**, 262.
- Keilin, D. 1926 *Proc. Roy. Soc. B*, **100**, 129.
- Kowalevsky, A. 1892 *Int. Congr. Zool. (Moscow)*, **1**, 187.
- Lemberg, R. 1937 *Perspectives in biochemistry*, p. 137. Cambridge.
- Lemberg, R., Legge, J. W. & Lockwood, W. H. 1939 *Biochem. J.* **33**, 754.
- Lemberg, R., Legge, J. W. & Lockwood, W. H. 1941 *Biochem. J.* **35**, 328.
- Lemberg, R. & Wyndham, R. A. 1936 *J. Roy. Soc. N.S.W.* **70**, 343.
- Lennox, F. G. 1940 *Pamphl. Coun. Sci. Industr. Res., Aust.*, **102**, 51.
- McNee, J. W. 1913 *J. Path. Bact.* **18**, 325.
- Nordenskiöld, E. 1908 *Zool. Jb. (Anat.)*, **25**, 637.
- Nuttall, G. H. F. & Keilin, D. 1921 *Parasitology*, **13**, 184.
- Patton, W. S. & Cragg, F. W. 1913 *A textbook of medical entomology*. Calcutta.
- Poisson, R. 1926 *Arch. Zool. Exp. gén.* **65**, 181.
- Poll, M. 1934 *Rec. Inst. zool. Torley-Rousseau*, **5**, 73.
- Radu, V. 1932 *Arch. Anat. micr.* **28**, 107.
- Rich, A. R. 1925 *Physiol. Rev.* **5**, 182.
- Samson, K. 1909 *Z. wiss. Zool.* **93**, 180.
- Storey, H. H. 1933 *Proc. Roy. Soc. B*, **113**, 463.
- Verzár, F. & McDougall, E. J. 1936 *Absorption from the intestine*. London.
- Waterhouse, D. F. 1940 *Pamphl. Coun. Sci. Industr. Res., Aust.*, **102**, 28.
- Wigglesworth, V. B. 1931 *J. Exp. Biol.* **8**, 411.
- Wigglesworth, V. B. 1933 *Quart. J. Micr. Sci.* **76**, 270.
- Wigglesworth, V. B. 1936 *Quart. J. Micr. Sci.* **79**, 91.
- Wigglesworth, V. B. 1937 *Biochem. J.* **31**, 1719.
- Wigglesworth, V. B. 1942 *J. Exp. Biol.* **19**, 58.
- Yuile, C. L. 1942 *Physiol. Rev.* **22**, 19.
- Zileh, A. 1936 *Z. wiss. Zool.* **148**, 89.

The spectral sensitivity of the fovea and extrafovea in the Purkinje range

BY H. V. WALTERS AND W. D. WRIGHT

Technical Optics Section, Imperial College of Science

(Communicated by H. Hartridge, F.R.S.—Received 17 December 1942)

The spectral sensitivity of the retina has been measured by finding the energy required at various wave-lengths through the visible spectrum to match the brightness of a fixed comparison field. Results are recorded for two observers for a series of brightness levels within the Purkinje range and for three retinal areas, the fovea, 3° from the fovea and 10° from the fovea. A small shift of the foveal luminosity curve towards shorter wave-lengths has been observed with decreased brightness of the matching field, accompanied by the development of a hump on the red side of the curve. This may be due to the progressive isolation of the red response at low intensities. For the extrafoveal areas, the usual Purkinje shift was observed, although surprisingly high intensities were required before the photopic curve was approached. This suggests that the rods may continue to function over a greater range of intensities than is generally accepted. No sudden change from rod to cone vision was recorded, in contrast to other visual functions, and of the possible mechanisms by which the transition from one system to the other may operate, the special tendency of the rod response towards convergence and summation at low intensities seems the most likely. Approximate solutions to problems in photometry concerned with the change in relative brightness of coloured surfaces with reduced illumination may be derived by comparison with the corresponding changes which occur between monochromatic radiations of the same hue.

THE SIGNIFICANCE OF THE LUMINOSITY CURVE

The relative sensitivity of the eye to different parts of the spectrum is an important function both in its bearing on visual theory and on photometric practice. The description of the spectral sensitivity usually takes the form of the 'luminosity curve of the equal-energy spectrum', and it is well known that this curve has two main forms, one corresponding to observations made at high intensities—the photopic curve, figure 1, curve (a)—the other corresponding to low intensity observations, yielding the scotopic curve, figure 1, curve (b). The scotopic curve is very similar to the absorption curve of visual purple, the photochemical substance that has been extracted from the rods, and it is generally assumed that the photopic curve bears a similar relation to the absorption curve of the photochemical substance or substances that are believed to exist in the cones.

The luminosity curve shows the relative brightness of different parts of the spectrum, but whether or not we believe that a meaning can be attached to the phrase 'magnitude of sensation', such that ratios can be expressed between unequal sensations, there is no agreed quantitative correlation between stimulus and sensation, hence any numerical values given to the brightness of different spectral regions do not refer to the sensations. The proper meaning to be attached to the luminosity values can be understood by considering the basic observation performed when they are measured. Thus when the relative brightness of two mono-

chromatic beams of wave-length λ_1 and λ_2 is to be determined, the beams are arranged to illuminate the two halves of a photometric field and the intensity of one is adjusted until it appears equal in brightness to the other. If, for this equality setting, the energies in the two beams are in the proportion of $e_1:e_2$ respectively, then the luminosities are said to be in the inverse ratio, namely, $e_2:e_1$. Hence equality in the magnitude of two sensations is the only subjective judgement required in measuring the luminosity curve, and the values themselves merely indicate the inverse of the relative energies necessary to secure this equality.

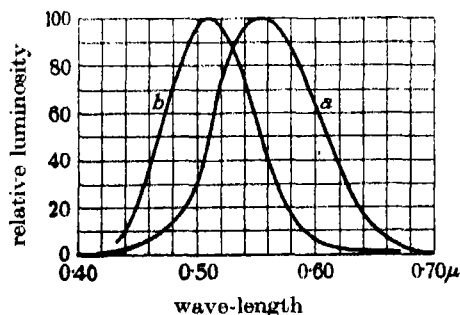


FIGURE 1. The luminosity curve of the eye (*a*) at high intensities, (*b*) at low intensities.

The calculations can be extended to any coloured beams of light beyond those restricted to narrow bands in the spectrum, by applying the law of the addition of luminosities. This law states that if two beams of light which are themselves equal in brightness are added respectively to two other equal beams, then the combined beams will also be equally bright. Thus if a beam of red light, luminosity L_R , were matched against a blue beam, luminosity L_B , so that

$$L_R = L_B,$$

and if a beam of green light, luminosity L_G , were matched against a yellow beam, luminosity L_Y , so that

$$L_G = L_Y,$$

then the luminosity of the mixed red and green beams should be equal to that of the blue and yellow beams together, giving

$$L_R + L_G = L_B + L_Y.$$

It should be noted, however, that the left-hand side of the equation now refers neither to red nor to green light but will have some different quality such as yellow, while the right-hand side may refer to a white. The law of addition of luminosities is an experimental law and only holds for observations made either at high brightness levels where the sensitivity curve of the eye is that given by the photopic luminosity curve or at such low brightnesses that the sensitivity corresponds to the scotopic luminosity curve. At intermediate levels the law breaks down.

When the relative brightness has to be found for two beams which are not monochromatic, it is necessary to sum up the luminosities of the spectral components in each beam, and compare the sums mathematically. If the quality of the beams is defined by the distribution of energy through the spectrum, the energy values will first have to be converted into luminosity values by multiplication with the luminosity factors taken from the equal-energy luminosity curve. The products must then be summated. But again, when the relative brightness has been found, it is not the relative magnitude of the sensations that has been determined but the amount by which the intensity of one beam must be increased or the other decreased in order that the two should appear to be equally bright.

Thus the luminosity curve, in addition to its relation to the absorption curve of the light-sensitive material in the retina, also expresses the relative brightnesses of all coloured objects in the field of view. Its position as one of the most important characteristics of vision hardly calls for emphasis, and so far as the photopic curve is concerned, a great deal of experimental work has been carried out in its determination. The same applies, but to a less degree, to the scotopic curve, and the need for further observations might be thought difficult to justify. There are, however, two aspects of the phenomena which appear to have received little attention: very few reliable measurements (König 1903; Sloan 1928, where an excellent and exhaustive account of previous investigations is given; Weaver 1937) have been made at brightness levels intermediate between the purely photopic and purely scotopic levels, and little is known regarding the spectral sensitivity of different areas of the retina. Without these two additional sets of results it cannot be said that our information is complete, because several questions of physiological interest would remain unanswered. For example: (1) Have the extrafoveal cones the same spectral sensitivity as the foveal cones? (2) How do the rod and cone responses combine when both are stimulated simultaneously? (3) Do the foveal cones show any change in sensitivity with change in brightness level? (4) Over what range of intensities is the scotopic curve virtually unaffected by changes in brightness? (5) Why is the rod curve suppressed at higher intensities? These questions are discussed in detail later in this paper, although it would be foolish to suggest that complete answers have been found. What cannot be gainsaid is that they are questions of importance, and that if answers are to be attempted, luminosity curves at various brightness levels and various retinal areas are required.

Furthermore, there are practical problems in photometry at low intensities that are not only difficult to solve but are difficult even to formulate unless we have a clear understanding of the way in which the eye functions at intermediate levels between the purely photopic and purely scotopic states. These problems cannot be ignored on the assumption that this intermediate range occupies only a narrow and insignificant band in the whole gamut of intensities. On the contrary, as will be apparent from later sections of this paper, the range of intensities over which vision is a compound function of both rods and cones is very extensive, disturbingly extensive from the point of view of photometry.

The experiments reported in this paper were planned to cover this intermediate range of intensities and to test the variation in sensitivity of different retinal areas. The results obtained have been analysed in relation to the problems outlined above, but owing to the length of the investigation, results have only been recorded for the two authors. It is hoped that this paper will represent the first instalment of a general investigation into the colour vision of the retinal area surrounding the fovea.

APPARATUS AND METHOD OF OBSERVATION

The colorimeter developed by Wright (1927, 1939) was used throughout the investigation. As adapted for luminosity measurements, it acted as a double monochromator in which the two halves of a photometric field, 2° square and divided horizontally, were illuminated by monochromatic pencils of wave-lengths λ_1 and λ_2 . The brightness of the lower half of the field (λ_1) could be adjusted to any desired value by inserting a neutral filter of the appropriate density, while that of the upper half (λ_2) could be varied continuously by means of a photometer wedge and also in steps by means of neutral filters. The intensity range could be extended still further with the aid of any one of a series of fixed aperture sectors, rotating to intercept both beams at a speed greater than the critical frequency of flicker.

With the arrangement as described, the task of the observer consisted merely in adjusting the intensity of the upper field until it appeared to be equal in brightness to the lower half under the given conditions of observation. This was repeated for different values of λ_2 through the spectrum. The two halves of the field in general differed in colour and, as is well known, the matching of the brightness attribute of the two patches under these circumstances is more uncertain than the usual photometric setting between fields of the same colour. As a rule, however, the observations were less difficult than had been anticipated, and at the lower intensities the colour was less obtrusive and correspondingly less disturbing to the accuracy of matching. The alternative to using the direct comparison method was to employ a flicker method, but for the majority of the observations the direct comparison method was preferred because it avoided uncertainty in the interpretation of results obtained by the flicker method.

For the measurement of the photopic curve itself, however, the following flicker method was employed. The observer's retina had presented to it first λ_2 and then a white surface; then λ_2 once more and then again white and so on in rapid alternation. The white illumination was provided by a rotating sector which interrupted the λ_2 beam, the front surface of the sector being coated with a layer of magnesium oxide and illuminated by white light. The white illumination was set to a suitable level, the speed of the sector adjusted to a critical value, and the intensity of the monochromatic beam varied until flicker disappeared. This was repeated for a series of values of λ_2 through the visible spectrum.

The exit-pupil of the colorimeter consists of a small circular aperture about 1 mm.

in diameter. With this constriction of the rays as they enter the eye, it is very desirable that the observer's head should be fixed relative to the instrument to ensure that the pencils of light pass through the centre of the eye pupil. This was arranged by using the dental impression mouthpiece fixed to the apparatus; the observer then bites on this to locate and fix his head in the required position.

When the sensitivity of the extrafovea was being tested, a small red fixation spot was introduced at the appropriate distance from the photometer field. Three retinal areas were tested in all, the fovea, a position 3° to the nasal side of the fovea, and a position 10° to the temporal side. These distances refer to the separation between the centre of the 2° field and the fixation spot, which was, of course, viewed foveally. For the foveal measurements, the photometer field was itself used as the fixation object. This presented no difficulty at the higher intensities, when the fovea is still the most sensitive part of the retina, but for the lower intensities, the viewing of the patch on the fovea presents very great difficulties owing to its reduced sensitivity relative to the parafovea. The difficulty can be overcome in two or three ways, the method adopted being to select the fixed comparison patch from the red end of the spectrum by making $\lambda_1 = 0.63\mu$. Since the fovea is populated very largely, if not entirely, by cones and the rods are insensitive to red light, the fovea remains the most sensitive area of the retina for this wave-length even when its intensity is made very low. Hence the eye naturally rotates to bring the patch on to the fovea and foveal measurements become possible. Even so, they could not be described as easy in those cases when the other half of the field was illuminated by green or blue light, for the latter tended to pull the patch outside the fovea. Great concentration on the part of the observer was necessary to ensure that the observations were made for the fovea, and the employment of the red comparison field served the double purpose of aiding the concentration and acting, through its increased brightness, as a signal that the patch was being observed foveally. Although a 2° field is slightly larger than the fovea, the retinal area covered by the patch can be regarded as effectively free of rods (Polyak 1941); the use of a smaller field was not favoured owing to the loss in accuracy of matching that would have resulted.

The brightness of the field of view was calculated from measurements on the illumination of a white screen which, when viewed through the eyepiece end of the colorimeter, matched the 0.63μ comparison field at one of the higher levels at which it was used. When allowance had been made for the size of exit-pupil and for reflexion losses in the eyepiece system, the retinal illumination in photons was derived. Other levels could be calculated from the density of the neutral filters and the opening of the sectors used to give each particular level, but at the lower levels the meaning of the values obtained when expressed in ordinary photometric units is by no means clear owing to the changing luminosity curve. The proper specification of the brightness levels is therefore considered in detail later.

A period of 20 min. dark adaptation was allowed before any observations commenced. The matching field was exposed continuously during the observations,

although in practice the field was observed only for short intervals while actually making a match. The readings of the photometer wedge were recorded by the person operating the apparatus, who was also responsible for setting the intensity level, controlling the voltage of the lamp used in the colorimeter and for changing the wave-length of λ_2 . Three matches were made at each wave-length and the mean of the three readings taken. The luminosity curve for only one intensity level was recorded at any one session. It was hoped that the curve for each condition of observation, i.e. intensity level and retinal area, would be measured on three separate occasions to give a mean curve, but while this was possible in many cases for W.D.W., there was insufficient time available for the whole programme to be completed by H.V.W. The observations extended in all over a period of about 12 months.

MEASUREMENT OF THE ENERGY DISTRIBUTION IN THE SPECTRUM

For the preliminary energy measurements, a Kipp and Zonen vacuum thermocouple was used in conjunction with a Paschen galvanometer. The light emerging from the exit-pupil of the colorimeter was focused by means of a microscope objective on to the junction of the thermocouple. Elaborate precautions were taken with regard to screening and heat insulation for both thermocouple and galvanometer, but in spite of these the random movements and steady drift of the galvanometer were considerable in comparison with the relatively small galvanometer deflexions that could be obtained. The energy measurements were made for the beam in the lower half of the field since more light could be obtained than in the other beam, and while satisfactory readings were recorded for wave-lengths between 0.70 and 0.50μ , the results below 0.50μ were less reliable, and for wave-lengths shorter than 0.47μ the values were obtained purely by extrapolation. It was unfortunate that measurements were not possible over the whole of the visible spectrum, although the most important region was covered. Even in those parts of the spectrum that were not properly measured, the comparison of the luminosity curves for different brightness levels and retinal areas is still of significance.

The energy distribution curve as obtained direct from the galvanometer deflexions applied, as stated above, to the lower field. In the visual observations, the illumination of this field was kept constantly at the wave-length of 0.63μ , while the wave-length of the light in the upper field was varied. The energy measurements had therefore to be transferred to the upper field, a question only of making a series of visual matches between corresponding wave-lengths in the two spectra and noting the readings on the photometer wedge. From the energy values and the calibration curve of the wedge, the wedge readings for uniform energy distribution through the spectrum were calculated. (Actually corrections had also to be applied to allow for the small departures from neutrality of the photometer wedge and, when neutral filters were interposed in the upper beam, of the filters used. These

corrections were obtained from subsidiary calibrations on the colorimeter.) From the energy measurements just described, the wedge readings for constant energy were found; from the luminosity matches themselves, the wedge readings for constant brightness were recorded. The ratio of the intensities corresponding to the two readings then yields the relative luminosity of different wave-lengths through the spectrum.

No allowance has been made in the results for the fact that the mid-wave-length of the light transmitted by the exit-pupil is not identical with the average wave-length when each elementary wave-length is weighted by its energy and luminosity. The band of wave-lengths transmitted by the exit-pupil varies from about 0.0020μ in the blue to 0.0060μ in the red, but as the difference between the arithmetic mean and the weighted mean would only amount to a very few angstroms, the error, relative to other observational errors, is insignificant and has been neglected.

RESULTS

The only point to be settled before giving the experimental results is the question of defining the intensity levels of the 0.63μ comparison patch. In ordinary photometric problems, the brightness of a field can be expressed in the usual units of candles per unit area or, if the interest in the problem is primarily visual, the retinal illumination in photons can be given. In the present experiments, photometric units have little or no meaning at the lower intensity levels. As already mentioned, the brightness of the red field was measured at a high intensity level and the result could have been expressed in terms of photons; they would have been red photons, but they would have been equal to the same number of green photons or photons of any other colour, and the value would have had a definite meaning. If, now, the intensity were reduced to $\frac{1}{100}$ th or $\frac{1}{1000}$ th of its previous value, the new retinal illumination might be calculated by reducing the number of photons by the same factor, but if the corresponding reduction were made in the hypothetical green or white photons, the three could no longer be equal in brightness although represented by the same numerical values. Hence we must either arbitrarily choose a colour in which to develop a scale of photons, for example, a scale based on a given white illumination, or else adopt some other method of specifying the brightness of the comparison field. Even if we established a scale of white photons, the problem would only have been partially solved, since the intensity of any other colour would have to be found by matching against the white, and the result of the match would vary according to the area of the retina on which the match was made. Hence while a white scale may be necessary and adequate for purposes of illuminating engineering, its meaning is not sufficiently precise to use here as a reference scale. For these reasons, it was decided that, to avoid all possible ambiguity, the intensity of the fixed comparison field of wave-length 0.63μ should be specified in terms of energy values, that is, in terms of the number of ergs falling on unit area of the retina per sec. Direct measurement of

the energy in absolute units was hardly feasible and was in any case unnecessary, since the brightness of the field under photopic conditions had been measured and the luminous efficiency of radiant energy is known (650 lumens per watt at a wave-length of 0.555μ ; Wensel 1939). From these values and from the luminosity of 0.63μ relative to 0.555μ , it was a straightforward matter to transform from photons to ergs. The transformation for 0.63μ radiation gives 1 photon as equivalent to 2.6×10^{-5} erg per sec. radiated in each square degree of the visual field. The energy might have been defined per unit area of the retina, but an angular measure of the field of view is equally valid and more directly related to the usual visual problems. The incident energy for the other intensity levels at which observations were made was then calculated directly from the transmission factors of the filters and sectors used to control the intensity. It should be recorded that the observations were made using a circular artificial pupil in front of the eye, the area of the pupil being 1.24 sq.mm.; it seems unlikely, however, that the measurements would be materially affected if the pupil had been larger provided the intensity of the light were reduced to keep the energy falling on the retina unchanged, except that, with a much larger pupil, the Stiles-Crawford effect (Stiles & Crawford 1933) would have to be taken into account in relating the intensity and pupil area to maintain constant effective energy.

(a) *Foveal luminosity curves.* At very low intensities approaching the foveal threshold, both observers on occasion obtained curves showing two maxima, one at a wave-length of about 0.60μ and the other at about 0.54μ . They have not been incorporated in the published results because they could not be recorded with consistency; subsequent investigation may show them to have been artefacts due to poor fixation or to the extension of the photometric field beyond the rod-free area of the retina. It would be unwise to give them a prominence which they did not merit, although, if confirmed, they would prove of considerable interest, especially in view of Granit's recent work (1941a) on the spectral sensitivity of individual receptors.

The remaining curves for the fovea are illustrated in figures 2 and 3.

Of the three curves given for H.V.W., the two lower intensity curves were obtained by direct comparison between the two halves of the field, while the highest intensity was his photopic curve as measured by flicker. The five curves given in figure 3 are for W.D.W., the curves for the four lowest intensities being recorded by direct comparison, the fifth and highest intensity by flicker. The curves have been numbered according to the intensity of the comparison field (wave-length 0.63μ) against which they were matched; the energy values to which these numbers correspond can be found by reference to table 1.

(b) *3° luminosity curves.* Figures 4 and 5 give the series of curves obtained by H.V.W. and W.D.W. respectively when the centre of the 2° field was located at a position 3° to the nasal side of the fovea. The curves are again numbered to correspond to the intensities given in table 1, and it will be seen that six curves are given for H.V.W. and seven for W.D.W. In addition, the foveal photopic curve is shown

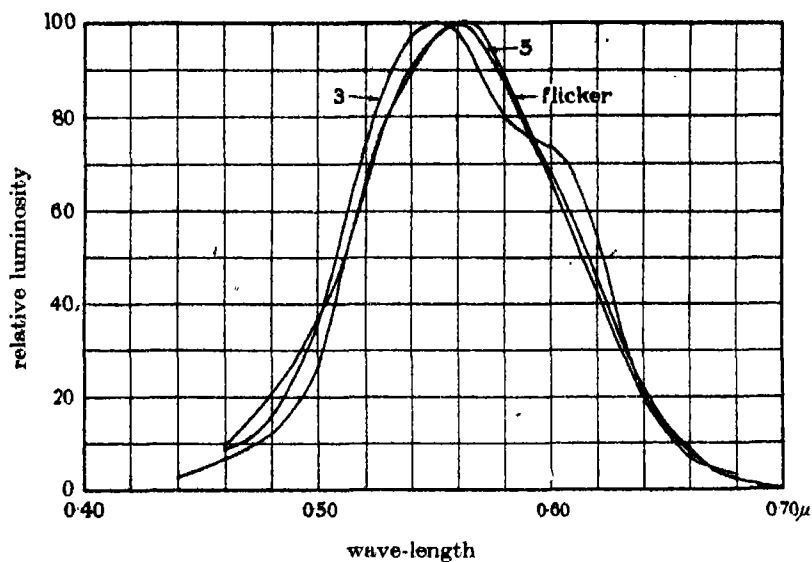


FIGURE 2. Luminosity curves for the fovea (H.V.W.). The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

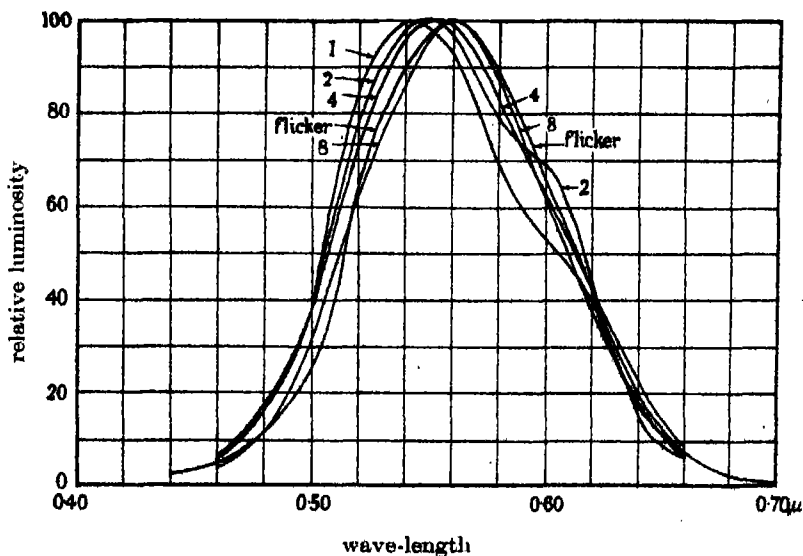


FIGURE 3. Luminosity curves for the fovea (W.D.W.). The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

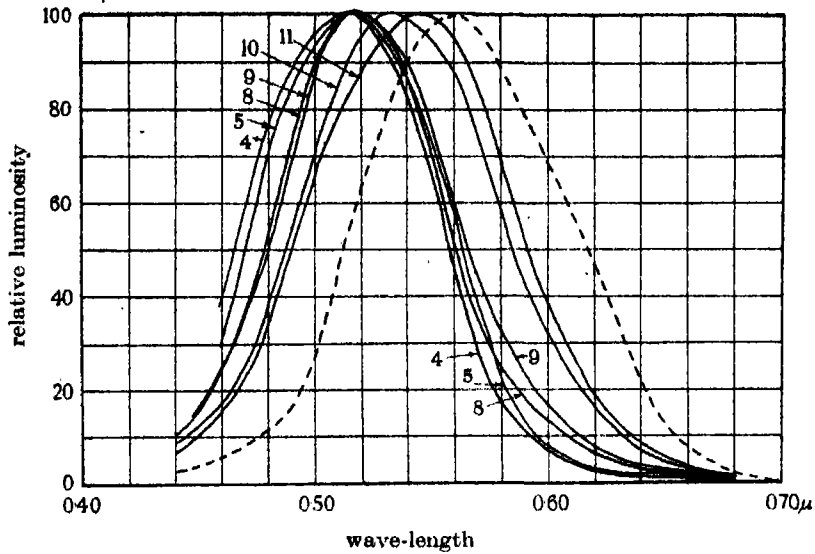


FIGURE 4. Luminosity curves for retinal area 3° from fovea (H.V.W.). Foveal photopic curve shown dotted. The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

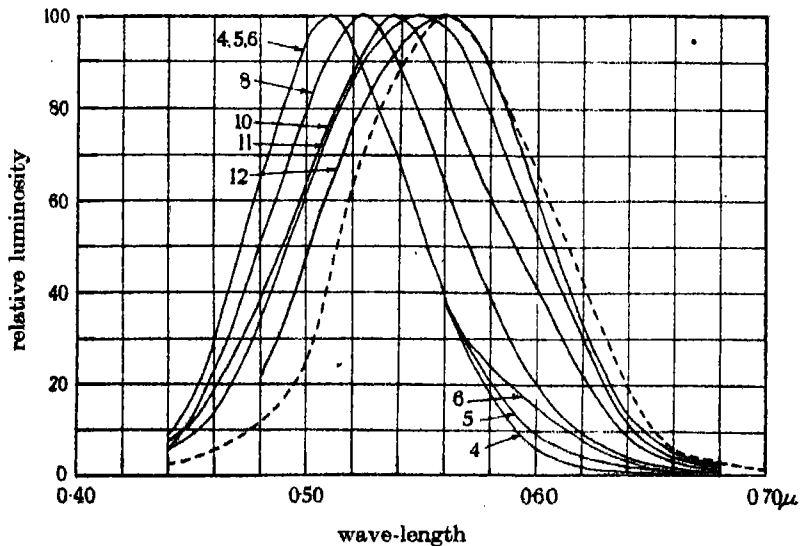


FIGURE 5. Luminosity curves for retinal area 3° from fovea (W.D.W.). Foveal photopic curve shown dotted. The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

TABLE 1. INTENSITY LEVELS AT WHICH LUMINOSITY CURVES HAVE BEEN MEASURED

level number	energy density of 0.63μ field (ergs/deg. ² /sec.)	level number	energy density of 0.63μ field (ergs/deg. ² /sec.)	level number	energy density of 0.63μ field (ergs/deg. ² /sec.)
1	0.015×10^{-4}	6	0.092×10^{-4}	11	0.93×10^{-4}
2	0.025	7	0.10	12	2.3
3	0.031	8	0.15	13	5.7
4	0.038	9	0.25	14	14
5	0.062	10	0.37	15	45

for comparison. The three lowest intensity curves for W.D.W. showed little systematic difference for wave-lengths shorter than 0.56μ , consequently the average of the three has been shown over this region. At the red end of the spectrum, the curves deviated to a significant extent and have therefore been drawn separately.

(c) *10° luminosity curves.* Figures 6 and 7 give the series of curves obtained by H.V.W. and W.D.W. respectively when the centre of the 2° field was located at a position 10° to the temporal side of the fovea. Four curves were recorded by H.V.W. and six by W.D.W., but in neither case could the curves for the highest intensity levels be extended below a wave-length of 0.48μ owing to the limited amount of light available at the blue end of the spectrum.

Accuracy. The accuracy varied over a wide range according to the type of experiment and also in a subsidiary fashion on the observer's ability to concentrate, so that only the broad outlines of the accuracy which may be claimed can be defined.

Since the wave-length of the comparison field was 0.63μ the matches were most consistent at the red end of the spectrum and least in the green and blue, and although the foveal observations might be expected to be more reliable than those made at 3° and 10° , this was only true for the higher intensities. At low intensities, fixation difficulties increased the errors of the foveal matches and as a rule the 10° curves were those most easily reproduced. At 3° , while the eye was still sensitive to comparatively small differences in intensity, the rapidly changing sensitivity of the retina with distance from the fovea led to relatively large errors when wandering of fixation occurred. On the whole, the results obtained by W.D.W. are more reliable than those by H.V.W., partly because the results were less scattered and were reproduced more consistently, and partly because more time was available for repeat observations by W.D.W. Generally, an accuracy of the order of $\pm 10\%$ can be assumed, with an additional uncertainty at wave-lengths shorter than 0.47μ , because the energy values were obtained by extrapolation.

Although the observer was dark-adapted before commencing the observations, the photometer field itself produces a local adaptation effect on the retina which will be greater the higher the intensity of the field. The quality of the adaptation

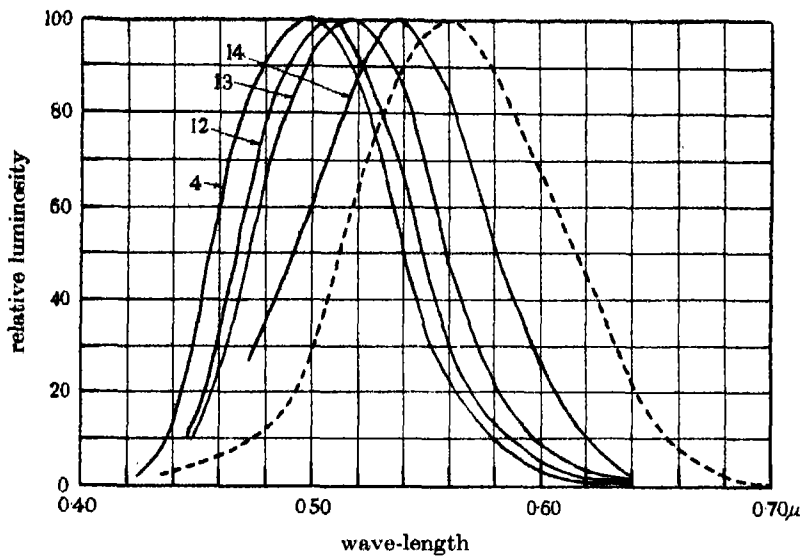


FIGURE 6. Luminosity curves for retinal area 10° from fovea (H.V.W.). Foveal photopic curve shown dotted. The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

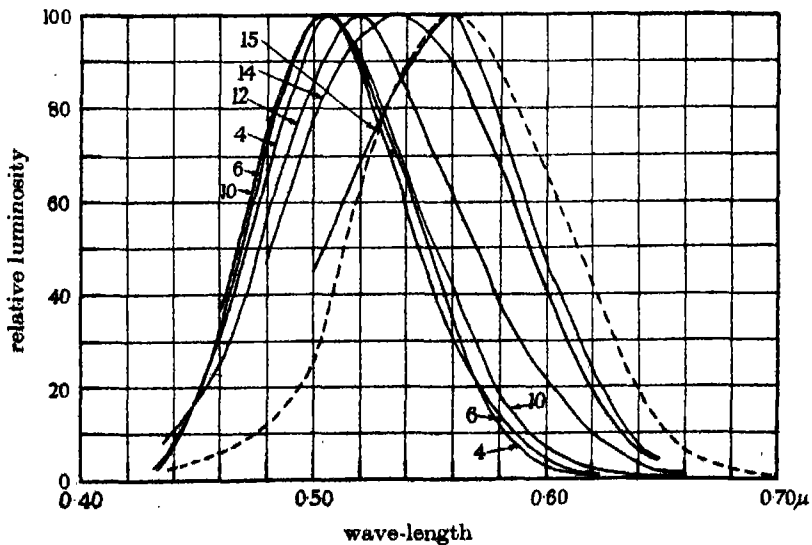


FIGURE 7. Luminosity curves for retinal area 10° from fovea (W.D.W.). Foveal photopic curve shown dotted. The reference numbers indicate the intensities at which the curves were measured, as given in table 1.

will depend on the colour of the field, a red patch depressing the red sensitivity the most and so on. The effect is difficult to avoid or allow for, but it can be reasonably assumed that at low intensities the adaptation is unimportant, and at no time during the matching was the field fixated for any lengthy period. Strictly, any recorded match can only be said to hold for the particular sensitivity of the retina at the time at which the match was made. Guild (1932) has maintained that because of the adaptation processes, the ordinary manipulations of luminosity equations are not legitimate; that if, for example, a green and blue patch were each in turn made equal to a red patch, it would not be correct to assume that the green and blue patches would match when compared side by side, since the conditions of observation would have changed. There can be no doubt that this is a fundamentally correct outlook, and if we do not adopt it in the following discussion, it is not through ignorance or disagreement, but because, in the absence of evidence to the contrary, we do not believe the magnitude of the discrepancy is sufficiently great to warrant so restricted an application of our experimental results.

DISCUSSION OF THE RESULTS

(a) *Fovea*. Although the changes in the spectral sensitivity of the fovea with reduction in intensity are small, they are systematic and significant. The tendency is for the red ordinates to decrease and the blue to increase with decreasing intensity of the matching field, the position of maximum sensitivity shifting about 0.02μ towards the shorter wave-lengths. Further, a definite hump is recorded on the red side of the curve, a hump that was also found by Sloan (1928).

The absence of any major shift of the curves is an expected consequence of the absence of rods from the fovea. The small shift towards the blue may be due to the small number of rods that may exist within a 2° area around the fovea or within the area covered by the inevitable wanderings from true foveal fixation. Whether or not this is the correct explanation, it is hardly likely that a combination of rod and cone sensitivity could produce a compound curve with a hump at about 0.60μ . It is more probable that the red receptors of the Young-Helmholtz theory tend to maintain their sensitivity over a greater intensity range than do the green and blue receptors, a suggestion that has been explored by Forbes (1929) on the basis of Sloan's measurements.

(b) *Area 3° from fovea*. In this area the change from the photopic to the scotopic curve is fully developed. With decreasing intensity the maximum has shifted from 0.56 to 0.51μ , but the position of the maximum is not necessarily the most striking or significant feature of the sensitivity curve. The general meaning of the change in the luminosity curve, e.g. the relative darkening of red colours, is too well known to need emphasis here, and only special features need be noted. One important point is that, for both observers, the red and orange luminosity values in the low intensity curves continue to diminish by important amounts even though the green and blue values show little or no change. There is no indication that the red

luminosity values have reached a steady level, and if observations could have been continued at still lower intensities, the relative luminosity of the red would have continued to diminish. The deduction must therefore be drawn that, from about 0.60μ and upwards, the sensitivity as recorded must correspond very largely to cone sensitivity; for although the rods may be completely insensitive in the red, a match will always be possible if enough light is admitted for the cones to respond.

Another aspect of the results is that the sensitivity is no longer affected by absorption of light in the yellow macular pigment. When the highest intensity curve is compared with the foveal photopic curve, a big difference on the blue side is apparent. In the case of H.V.W., this difference is only in part due to absorption by the yellow pigment; some of the difference arises because the intensity was not sufficiently high to give the true cone response, as shown by the quite large difference still outstanding on the red side. But in the case of W.D.W., the highest intensity 3° curve approaches quite closely to the foveal photopic curve on the red side, yet differs considerably on the blue side. At 0.50μ , for instance, it would appear that the macular pigment absorbs about 50 % of the incident light.

In so far as the subjective impressions of the observations are of interest, it should be mentioned that the matching field for the lowest intensity curve appeared very faint indeed, although no measurements were taken to determine its magnitude in relation to the absolute threshold. For the three or four lowest intensity curves, little sensation of colour was aroused by the blue and green wave-lengths, but the redness of the longer wave-lengths could always be recognized. For the highest intensity curves, the colours throughout the spectrum were easily visible and the field was sufficiently bright to give the impression that the cones alone would be involved. The curves themselves show that this was not the case and that the Purkinje phenomenon undoubtedly persists at higher intensities than the authors had expected. These remarks apply equally to the 10° curves.

Reference to the energy values in table 1 suggests that the main change-over from scotopic to photopic vision occupied a comparatively narrow range of intensities. This is deceptive, however, and is due to the fact that the energy values refer to the red comparison field. Owing to the changing luminosity curve, a 10 to 1 reduction in red intensity is equivalent to a much greater reduction of green light or white light. The red intensity scale is therefore more condensed than a corresponding green or white scale and allowance must be made for this fact.

(c) *Area 10° from fovea.* The results for this area are of a character generally similar to those of the 3° area. A large shift of the luminosity curve towards the blue with decreasing intensity is revealed and the red sensitivity at 0.60μ and beyond continues to change after the blue and green parts of the curve have become practically stable, more especially for W.D.W. The highest intensity curves show a higher sensitivity in the blue than the foveal photopic curve owing in part to the absence of the yellow macular pigment. But for neither observer does the intensity appear to have reached a sufficiently high value for it to be likely that the pure cone luminosity curve has been recorded in this area. It would

therefore be unwise to conclude that the extrafoveal cones possess a different spectral sensitivity from the foveal cones, although it is surprising that a closer approach was not reached between the two curves at the intensities that were used, since the highest intensity level (level no. 15 in table 1) is equivalent to 173 photons by foveal observation.

(d) *Scotopic luminosity curve.* The sensitivity curve for the rods alone is of both theoretical and practical interest, and an attempt has been made to derive an average curve from the lowest intensity curves reported here. The 10° curves for H.V.W. were not included as they were only recorded once, and their shape could not be guaranteed. A mean curve has been calculated from the lowest intensity

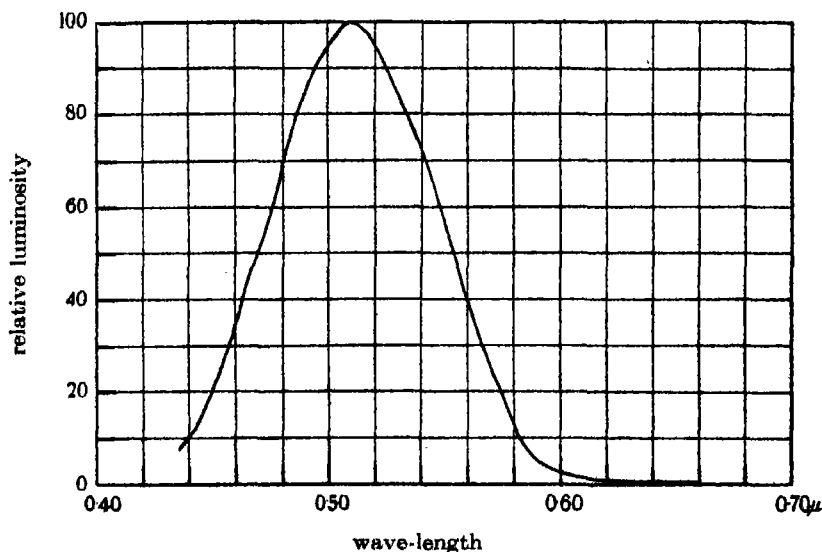


FIGURE 8. Mean scotopic luminosity curve from data for H.V.W. and W.D.W.

curves in each of figures 4, 5 and 7 and from a low-intensity curve recorded by H.V.W. in the early stages of the investigation before a fixation spot had been provided. From the blue end of the spectrum up to 0.56μ , the curves were all sufficiently similar to justify the assumption that they were all measures of the same function, namely, the rod sensitivity curve. An arithmetic mean of the curves up to 0.56μ was therefore calculated. But beyond 0.56μ , as has been mentioned above, there is strong evidence to suggest that as recorded they include an important fraction of the cone response, and a mean in this region is unlikely to give the closest possible approach to the pure scotopic curve. As the number of rods are known to increase relative to the number of cones with increasing distance from the fovea, it is likely that the 10° curves are more nearly scotopic than the 3° curves. The results themselves support this to the extent that the ordinates from 0.58μ onwards of the lowest intensity curve in figure 7 are noticeably smaller than the corresponding values in figure 5. It was concluded that in this region of the

spectrum the closest approach to the scotopic curve is accordingly given by the lowest intensity curve in figure 7, and the resultant scotopic curve, obtained by combining the mean results of H.V.W. and W.D.W. from the blue to 0.56μ with the single 10° curve of W.D.W. from 0.56μ to the red, is shown in figure 8.

For comparison, the most reliable curves published, as recorded by Hecht & Williams (1922), Sloan (1928) and Weaver (1937), are shown in figure 9. There is generally good agreement between the four curves, the most marked divergence, more noticeable if the luminosity values are plotted logarithmically as in figure 10, occurring in the red, as might be expected in view of the difficulty in eliminating all cone response. The authors' curve is rather low at 0.44 and 0.46μ compared to the other measurements, and this is probably due to the uncertainty in the energy distribution determination which, as reported earlier, was obtained by extrapolation below 0.47μ . It is, however, worth noting that the spread of the four curves in the blue end of the spectrum is surprisingly small compared with the enormous spread reported in investigations of the photopic curve (see, for example, Coblentz & Emerson 1918). The latter must presumably be caused by differences in density of the macular pigment from one observer to another, but even allowing for this, is extremely large.

The question arises as to the range of intensities for which the scotopic curve can be said to represent the luminosity curve of the eye. If the whole of the visible spectrum is included, then on the basis of the present results it appears that there is no range of any extent over which the scotopic luminosity curve is constant. In one sense, however, this is an incomplete statement of the facts. Thus, if we considered a light source with some fixed energy distribution—for convenience we might assume an equal-energy spectrum—and imagined the intensity to be gradually increased from below the visual threshold, then for some considerable range after the source became visible the response from the red end of the spectrum would remain subliminal, or at least contribute insignificantly to the total response. The luminosity ordinates in this region can then be ignored, and for the remainder of the spectrum the luminosity curve does appear to remain reasonably constant up to the intensity level 6 of table 1.

(e) *Relation between the rod and cone processes.* The change in visual response from the scotopic to the photopic function with increase in intensity of the stimulus cannot be explained on a simple hypothesis of the rod process responding at low intensities and the cones at high. It seems to be generally agreed that the initial reaction in the retina for both rods and cones is photochemical, and if Lambert's law holds as in ordinary photochemical reactions the fraction of light absorbed will be independent of the intensity of the incident light. When a beam of very low intensity strikes the retina only the rod process will react, because of the admitted low sensitivity of the cones and the existence of a finite threshold of stimulation; at higher intensities the cone threshold will be passed and both rods and cones will be responding simultaneously. The overall spectral sensitivity curve will then be some composite curve built up from the rod and cone components. Suppose

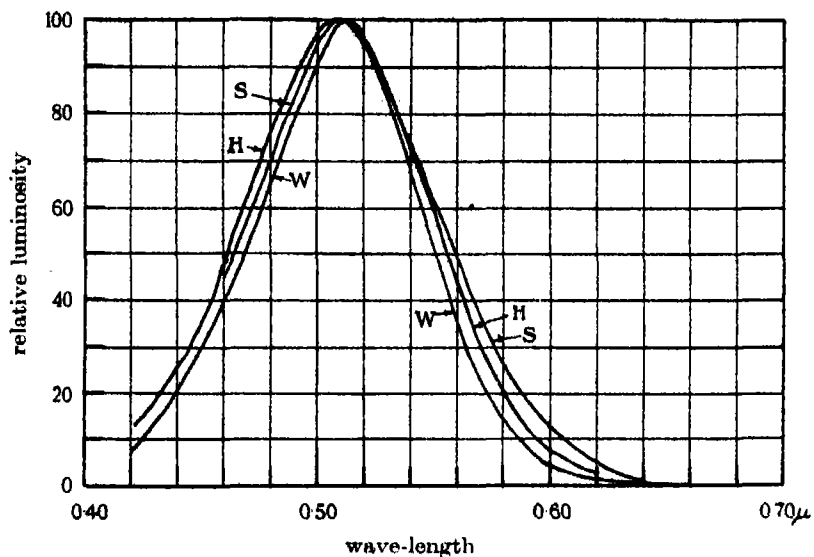


FIGURE 9. Scotopic luminosity curves as derived by Hecht & Williams (H.), Sloan (S.) and Weaver (W.).

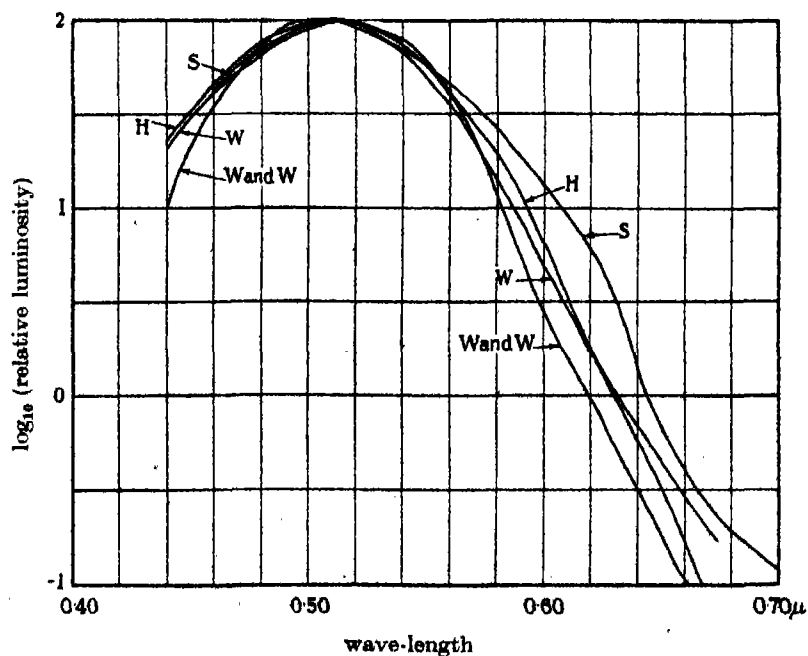


FIGURE 10. Scotopic luminosity curves of Hecht & Williams (H.), Sloan (S.), Weaver (W.) and the authors (W. & W.), plotted logarithmically.

now the intensity is increased still further; Lambert's law implies that the absorbed light will be increased by the same factor for both processes, but experiment shows that even if this is so, the rod and cone responses do not combine in some fixed proportion to yield a stable composite spectral sensitivity curve; the luminosity curve changes continuously until it passes from the scotopic to the photopic state.

Since the foveal cones show little tendency to develop a rod-like spectral sensitivity curve at low intensities, there appears to be no reason to suppose that the cone photochemical substance can transform itself into the rod substance after some extended period in the dark. Indeed, the very different functions and performance of the rods and cones provide strong support for the conception of two distinct processes, in the two kinds of receptor. Moreover, other evidence, quite apart from the spectral sensitivity curves, indicates that the cones are out of action at low intensities and the rods are inoperative at high intensities. Thus the light sensitivity of the retina is a maximum at the fovea at high intensities and falls off steadily towards the periphery, a result that can be reasonably explained on the known reduction in the number of cones with increasing distance from the fovea if we assume that the rods are inactive. Whether a close quantitative correlation could be established between the light sensitivity and the number of cones per unit area is by no means certain, but if the rods were acting, then there would be no reason to expect any falling off in sensitivity towards the periphery, but rather the reverse. On the other hand, at low intensities the position of maximum sensitivity occurs some 10° or so from the fovea and the overall variation in sensitivity across the retina is very much less than before, if the fovea, which is almost completely insensitive, is not included. This is in fair, although not exact, agreement with the distribution of the light receptors, if we assume that the cones do not function at low intensities.

We can conclude with some confidence that the rods and cones are actually different kinds of receptor and that the sensitivity within the overlapping range of intensities has to be explained in terms of their individual sensitivities acting together, and not in terms of a photochemical substance in a state of transition from rod to cone substance. The change from cone to rod sensitivity and vice versa may in part at least be connected with summation and facilitation effects by which, at low intensities and over small areas, the responses from weak stimuli tend to converge together to produce a resultant response which is greater than the sum of its parts. There is ample evidence (Polyak 1941) to show that the cones, through their more frequent association with midsize bipolar and ganglion cells, initiate responses which are less subject to summation than the rods, since the latter are in contact with the polysynaptic bipolar and ganglion cells. As the degree of summation that occurs is known to increase with decreasing intensity, then it follows that the rod response will diminish at a slower rate relative to the cone response as the intensity of the light stimulus is reduced. There need be no surprise, therefore, to find that the rod response tends to overtake the cone response at low intensities, but in the absence of more specific information about the processes of

facilitation and inhibition at the synapses, it is impossible to decide whether such action alone would account for the steady progression from the photopic to the scotopic luminosity curve, or whether some further inhibitory effect, blocking of the nerve paths or some other intermediate reaction, may be involved (Granit & Svaetichin 1939; Lythgoe 1940; Granit 1941*b*).

In any event, the constant brightness luminosity curves which we have recorded cannot be composite luminosity curves built up from the summed rod and cone curves taken in some given proportion. When a test stimulus from the red end of the spectrum is used and matched against the fixed comparison field, the response is derived mainly from the cones; when the stimulus is changed to, say, the green, the response may be a mixture of rod and cone impulses in roughly equal proportions, while with a blue stimulus the response will be primarily from the rods. Hence it is not possible to define the relative magnitude of the rod and cone components in a combined response merely from a knowledge of the brightness of the stimulus, since the ratio will vary with the wave-length of the stimulus. As nothing is known about the interaction between the rod and cone responses, it would therefore appear that there are at the moment too many unknown factors present to enable any deductions to be drawn from figures 4-7 as to the rate at which the rod response increases relative to the cone response with decreasing intensity.

Two general comments on the results may be made from a qualitative inspection of the curves. The first is that there does not appear to be any one intensity level at which the response switches over suddenly from being mainly cone response to mainly rod response, in contrast to other visual functions, such as dark adaptation, flicker, intensity discrimination, etc., in which sudden breaks have sometimes been recorded and which are generally interpreted as the transition point between the rod and cone processes (Hecht 1937). The second point is that although the curves for the 10° position have approached the foveal photopic luminosity curve and at still higher intensities would very probably reach it, yet the fact remains that at the quite high intensities that were used there still appears to be a considerable rod element in the combined response. It may therefore be that the activity of the rods persists to higher intensity levels than is generally imagined; the only other explanation would be that the peripheral cones have a somewhat different spectral sensitivity from the foveal cones.

(*f*) *Application of the results to photometric problems.* In the opening section of this paper, reference was made to the problem of comparing the brightness of two beams of light which radiate over an extended range of the spectrum. The general procedure is to multiply the energy at each wave-length in the two beams by the appropriate luminosity factor and compare the sums of the products within the limits of the visible spectrum. This method is only legitimate provided the comparison is made under conditions for which the luminosity curve is stable, i.e. unaffected by change of intensity, yet it may happen in practice that the relative brightness of two beams may be required when the brightness is within the Purkinje range. For example, if red and green diffuse reflecting surfaces were illumin-

ated by beams of white light, then at a certain brightness level the amount of light on the red surface might have to be k times greater than the amount of light on the green surface for the two to look equally bright. If the level of illumination is low, k will be a function both of the spectral quality of the light reflected from the two surfaces and of the brightness level at which the comparison is made. If we were dealing with monochromatic beams, the problem of finding k could be solved directly by reference to figures 4-7. The question now arises whether the results can be applied to beams that are not monochromatic.

From the discussion in the previous section, it is quite certain that we cannot employ one of the intermediate luminosity curves appropriate to the brightness level concerned, and apply the luminosity factors as if we were dealing with photopic calculations. That would merely ignore any possible interaction between rod and cone processes. Yet considering that the curves in figures 4-7 contain an extensive record of the changing sensitivity of the retina, it would be rather remarkable if at least an approximate solution to the problem, adequate for engineering purposes, could not be derived.

In an attempt to find such a solution, it has been assumed that the observing conditions are such that the results recorded for the 3° position would apply; accordingly, the curves in figure 5 have been replotted to give the curves shown in figure 11. Each of these curves refers to a single wave-length λ and shows the energy, plotted logarithmically, required at that wave-length to give a brightness match with a beam of wave-length 0.53μ for various energy values of 0.53μ . 0.53μ has been chosen as the reference wave-length, since it is roughly midway between the maxima of the photopic and scotopic curves and therefore the least susceptible of any radiation to a distortion of its intensity scale due to the Purkinje effect.

The curve in figure 11 for 0.53μ itself is, of course, a straight line inclined at 45° to the axis. When the luminosity curve is constant, e.g. in the photopic range of intensities, the curves for all other wave-lengths will also be straight lines, parallel to the 0.53μ curve and above or below it according as the wave-length is less or more luminous than 0.53μ . Thus at the higher energy values in figure 11, all the curves approach a slope of 45° , the curve for 0.56μ being slightly below that for 0.53μ , while those for 0.48 , 0.50 , 0.60 , 0.63 and 0.70μ are above the 0.53μ curve. At intensities sufficiently low for rods alone to be involved, we should again expect the curves to be straight and inclined at 45° , and while this result is actually obtained with wave-lengths up to 0.56μ , the curves for 0.60 , 0.63 and 0.70μ continue to diverge from the curve for 0.53μ . This provides a convincing demonstration that the ratio of the luminosities of these wave-lengths does not tend to a constant value and proves that cone processes are still involved. Intermediately between the high and low energy values, all the curves depart from parallelism to the 0.53μ curve and in some cases may cross it.

To understand how figure 11 might be used to solve the practical problem of the relative illuminations required on red and green surfaces to produce equality of

brightness at some given energy level, consider the curves for 0.53μ and 0.63μ . At an energy level of 6.0 log units (see figure 11), the ordinate for the 0.63μ curve is higher than the 0.53μ curve by about 0.55 log units, that is to say, 3.5 times the energy is required in the 0.63μ beam for it to be as bright as the 0.53μ beam. At an energy level of, say, 3.0 log units, the difference in ordinates has increased to 1.52 log units, so that the energy ratio for equal brightness has increased to 33.1 . The energy required in the 0.63μ beam has therefore increased relatively by a factor of 9.3 . If we could assume that the red and green surfaces behaved exactly as the 0.63μ and 0.53μ radiations, the problem would be solved. Either from direct observation or from a knowledge of their spectral reflexion characteristics, the

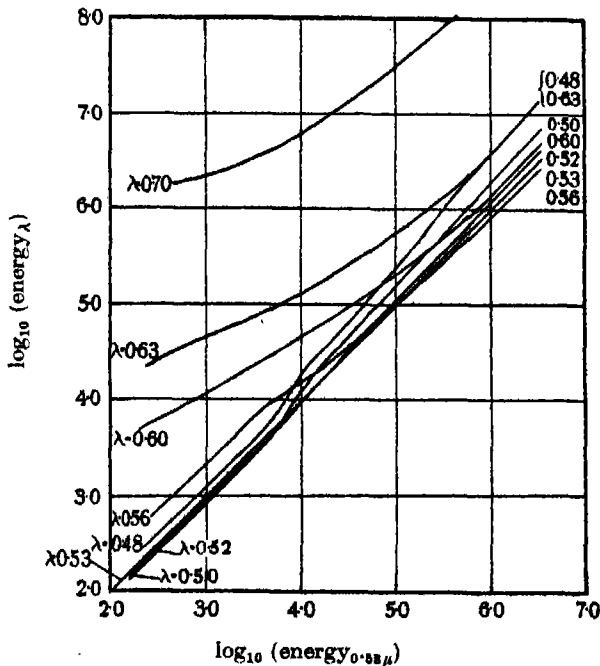


FIGURE 11. Curves showing energy required at wave-length λ to give brightness match with wave-length 0.53μ for various energy values of the 0.53μ beam. (Values calculated from results for W.D.W. for retinal area 3° from fovea.) Energy in $\text{ergs/deg.}^2/\text{sec.}$

relative illumination required for equality of brightness under photopic conditions could be found; the increase in this ratio with decreasing illumination could then be deduced by analogy with the monochromatic red and green beams. In practice, the pair of spectral colours and the pair of coloured surfaces will not behave identically. The more saturated the colours of the surfaces, the more closely will they follow the curves for the monochromatic beams. With desaturated colours and purples, the correlation with the behaviour of spectral colours will be less direct, although at the same time the effect of the Purkinje shift will be less marked.

The degree of approximation that could be tolerated in the solution of any

particular problem can only be decided by the nature of the problem, but figure 11 is likely to be of most use as an indication of the general trend and magnitude of the changes in relative brightness that occur between different colours. If in any given practical problem, coloured surfaces were required which changed in brightness in a particular manner or, alternatively, which avoided such changes, then figure 11 could be used as a guide to the selection of the appropriate colours. When any pair of surfaces had been selected, it would be no very difficult matter to make direct measurements on them at various illuminations to obtain the exact curve showing the change in their relative brightness.

While the solution of such problems may not present great difficulty when the area of the surfaces involved is fairly large, some rather awkward questions arise when the areas are small or when point sources are involved. Uncertainty is immediately introduced as to the area of the retina that is being used, and there is a consequent uncertainty in the luminosity curves that apply. As the illumination is lowered, each area of the retina changes from photopic to scotopic sensitivity, but further than that, the relative sensitivity of different retinal areas also changes, especially that of the fovea relative to the parafovea. At what level vision changes from foveal to parafoveal vision, whether the transition is sudden or gradual, whether, with small sources, the fovea is used for one colour and the parafovea for another, and how the change can be taken into account in the solution of practical photometric problems, are questions that we make no attempt to answer. To raise fresh problems, however, is no new event in research work and is welcome evidence that the subject is a live one and worthy of further study.

We wish to acknowledge with gratitude the continued financial assistance provided by the Medical Research Council for research on vision at the Imperial College.

REFERENCES

- Coblentz, W. W. & Emerson, W. B. 1918 *Bull. U.S. Bur. Stand.* **14**, 167.
 Forbes, W. T. M. 1929 *Amer. J. Psychol.* **41**, 517.
 Granit, R. 1941a *J. Opt. Soc. Amer.* **31**, 570.
 Granit, R. 1941b *Acta Physiol. Scand.* **2**, 98.
 Granit, R. & Svaetichin, G. 1939 *Uppsala LäkFören Förh.* (2), **45**, 161.
 Guild, J. 1932 *Reports on discussion on vision*, p. 60. London: The Physical Society.
 Hecht, S. 1937 *Physiol. Rev.* **17**, 239.
 Hecht, S. & Williams, R. E. 1922 *J. Gen. Physiol.* **5**, 1.
 König, A. 1903 *Gesamm. Abh. Physiol. Optik (Leipzig)*, p. 144.
 Lythgoe, R. J. 1940 *Brit. J. Ophthalm.* **24**, 21.
 Polyak, S. L. 1941 *The retina*. Chicago: The University of Chicago.
 Sloan, L. L. 1928 *Psychol. Monogr.* **38**.
 Stiles, W. S. & Crawford, B. H. 1933 *Proc. Roy. Soc. B*, **112**, 428.
 Weaver, K. S. 1937 *J. Opt. Soc. Amer.* **27**, 36.
 Wensel, H. T. 1939 *Bur. Stand. J. Res., Wash.*, **22**, 375.
 Wright, W. D. 1927 *Trans. Opt. Soc.* **29**, 225.
 Wright, W. D. 1939 *J. Soc. Instrum.* **16**, 10.

The visceral nervous system of the earthworm

II. Evidence of-chemical transmission and the action of sympatheticomimetic and parasympatheticomimetic drugs on the tone of the alimentary canal

BY FLIGHT-LIEUTENANT NORMAN MILLOTT, M.Sc., R.A.F.V.R.

*Lecturer in Zoology, The Victoria University, Manchester,
formerly Rouse Ball Student, Trinity College, Cambridge*

From the Zoological Departments, The Universities of Cambridge and Manchester

*(Communicated by H. Graham Cannon, F.R.S.—Received 25 January 1943—
Revised 26 June 1943)*

The action of adrenaline, ephedrine, tyramine and acetylcholine on the tone of the alimentary canal of the earthworm is considered in relation to the effects of nerve stimulation described in a previous paper. The similarity between the effect of the drugs and nerve action indicates that the antagonistic groups of nerves which indirectly influence the tone of the gut are probably cholinergic and 'adrenergic'. The indication of cholinergic nerves is confirmed by further, more direct, experimental evidence. The remarkable similarity which has been shown to exist between the type of indirect control of the gut in the earthworm and vertebrates is thus extended. The results of the experiments are considered in relation to those described by workers on other Invertebrates, and significant indications of a type of chemical transmission in earthworm remarkable in its similarity to that seen in vertebrates are described.

INTRODUCTION

The evidence presented in a preceding communication (1943) has shown that the alimentary canal of the earthworm, although exhibiting autonomous movements, apparently controlled by nerve elements in the gut wall, is indirectly controlled by two groups of visceral nerves which arise from the central nervous system and exert antagonistic effects on the tone of the alimentary musculature. In its broad outlines, therefore, the system of control is strikingly similar to that of the vertebrate gut, which is directly controlled by nerve plexuses within its walls, and, indirectly, by sympathetic and parasympathetic nerves exerting antagonistic effects.

It has long been known that many of the effects of sympathetic and parasympathetic stimulation can be faithfully reproduced by drugs, notably adrenaline and acetylcholine. Such observations form the basis on which the theory of chemical transmission in vertebrates was developed.

The action of acetylcholine and adrenaline in annelids has already been described by many workers, notably Gaskell (1914), Wells (1939 *a, b*), and Wu (1939 *a, b*). The work of the last is particularly interesting, since it deals with the response of the excised earthworm gut to various concentrations of acetylcholine and

adrenaline. These responses are not considered, however, in relation to the antagonistic responses evoked by stimulation of the segmental nerves under the conditions I described in a previous paper (1943).

In seeking evidence of chemical transmission in earthworm, it is therefore appropriate to commence by considering the effect of sympatheticomimetic and parasympatheticomimetic drugs on the tone of the gut, in relation to the effects of nerve stimulation described in a previous paper. An attempt will then be made to substantiate the conclusions drawn from these experiments by further, more direct, experiments.

THE ACTION OF SYMPATHETICOMIMETIC DRUGS—ADRENALINE, EPHEDRINE AND TYRAMINE

Methods

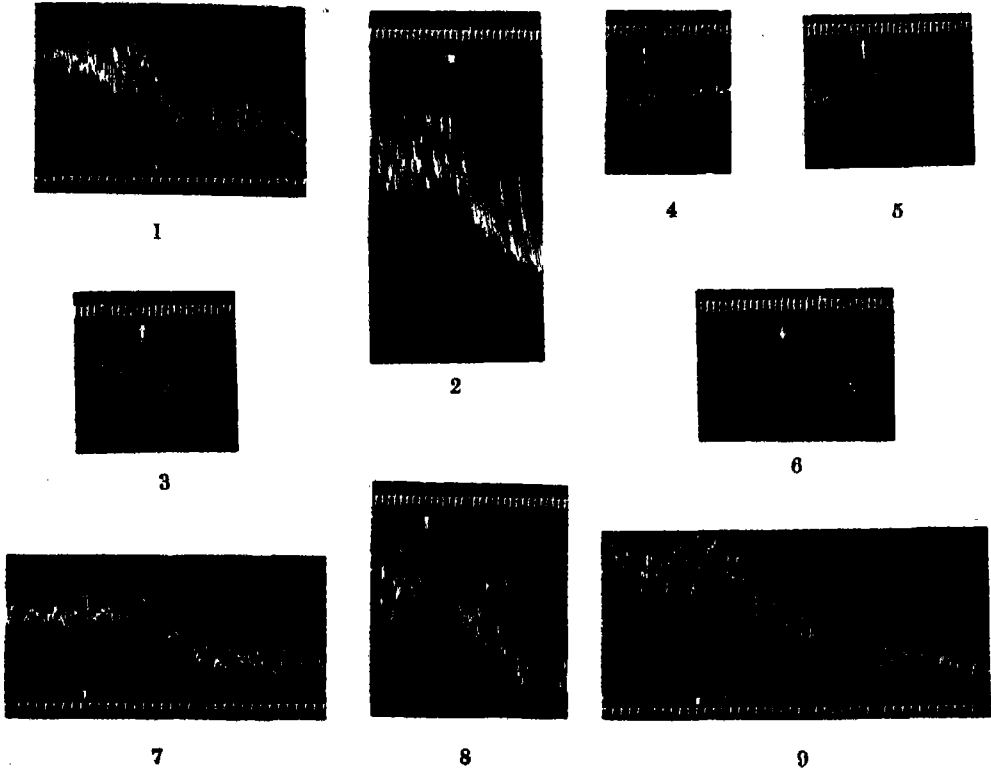
The earthworm, *Lumbricus terrestris*, was chiefly used in the following experiments. *Allolobophora longa* was sometimes used, but no significant difference in behaviour was observable.

Worms were opened up in the mid-dorsal line, and in each series of experiments one of the various regions of the gut was attached to the recording lever. About 0.02 c.c. of a solution of the drug in Ringer was injected into one of the hearts. The injection was made with a hypodermic syringe fitted with a glass nozzle, prepared by drawing out to an exceedingly fine point, a short length of 3 mm. diameter glass tubing, the final stages of drawing being performed over a match flame. The glass tubing was attached to the body of the hypodermic syringe by a piece of rubber tubing, the length of which was varied according to the disposition of the apparatus, but it was always long enough to permit the body of the syringe to rest on the bench, leaving both hands free to manipulate the nozzle. This was inserted into one of the hearts of a worm opened up under a binocular microscope.

The solutions used were prepared in the following ways. The adrenaline solution was made up from the solid 'Adrenalina' B.P. of British Drug Houses, Ltd., or 'Soloid' of Messrs Burroughs Wellcome and Co.; the former was dissolved in ordinary Ringer, the latter in Ringer buffered to pH 6.64 with Sorensen buffer (the amount of buffer used was such that the injected solution had only 1 % more sodium ions than ordinary Ringer). Solutions of ephedrine and tyramine were prepared by dissolving the solids in Ringer.

The effect on various regions of the gut of injecting about 0.02 c.c. of 1:1,000,000 solution of adrenaline is shown in figures 1, 2, 5 and 6. The effect of injecting the same amount of a 1:5000 solution of ephedrine on the tone of the oesophagus and crop is shown in figures 7 and 8, and the effect of injecting 1:13,000 tyramine on the oesophagus in figure 9. The injection of adrenaline (indicated by signal) was followed by a fall in tone of the oesophagus, gizzard and intestine, but the crop sometimes responded with a distinct rise in tone, though more often with a fall (figure 2). Thus, in a series of twelve experiments, the crop responded to the injec-

tion of adrenaline with a rise in four instances, and by a fall in the remaining eight. Since all the experiments were performed under precisely the same conditions, it is not possible to suggest any experimental conditions which might have favoured a rise or fall in tone in each instance.



FIGURES 1-9. Response of various regions of the gut to sympathetico-mimetic drugs injected into one of the hearts at signal

Time scale (as in succeeding figures) in half minutes.

FIGURE 1. Oesophagus: 1:1,000,000 adrenaline

FIGURE 2. Crop: 1:1,000,000 adrenaline.

FIGURES 3, 4. Controls, with crop and gizzard (respectively), Ringer injected.

FIGURE 5. Gizzard: 1:1,000,000 adrenaline.

FIGURE 6. Intestine: 1:1,000,000 adrenaline.

FIGURE 7. Oesophagus: 1:5000 ephedrine.

FIGURE 8. Crop: 1:5000 ephedrine.

FIGURE 9. Oesophagus: 1:13,000 tyramine.

Injection of ephedrine or tyramine is followed by a fall in tone, but the effect of the latter drug was studied on the oesophagus alone, that of the former on the oesophagus and crop.

Since the drugs were injected into the vascular system of a whole worm, they could become widely distributed, and therefore it could be argued that the fall in tone which is registered need not necessarily be due to a change in tone of the gut musculature, but might be due to the response of other effectors such as the muscles of the body wall and septa. In order to check this possibility, a series of control experiments was performed, in which the diameter of the oesophagus was measured before and after the injection of adrenaline. Worms were opened up under a binocular microscope fitted with a squared eyepiece, the body wall was firmly pinned out, the septa being stretched, but to avoid damage to the blood system of this region they were not cut. The results of three such experiments are given in table 1, the measurements being expressed in arbitrary units.

TABLE 1

experiment	diameter of oesophagus in arbitrary units	
	before injection	after injection
1	30	40
2	25	36
3	30	50

The possibility of the increase in diameter being due to an outward pull exerted by the septal muscles on either side is eliminated by direct observation, for if the gut is examined during and after injection, it is seen that the increase in diameter is by no means more pronounced at the point of insertion of the septal muscles, but on the contrary it is the intervening portions which dilate more actively. Such an effect would not be observed were the septal muscles responsible for the effect.

Thus adrenaline brings about a true inhibition of the gut.

Since Smallwood (1923) believes that the blood vessels are accompanied by a fine nerve plexus, it is possible that the process of injection may stimulate these nerves. However, control experiments show that such stimulation is not the cause of the fall in tone of the gut. Two such experiments are shown in figures 3 and 4. The solution of the drug was replaced by Ringer, which was injected under exactly the same conditions as in the foregoing experiments. The injection did not bring about any significant change in tone, and thus the change in tone observed in the foregoing experiments must be due to the drugs.

Wu (1939*a*) obtained different results with adrenaline on the earthworm gut. He states that the oesophagus responds to adrenaline, even in concentrations as low as 1:10,000,000, by a rise in tone. The crop, gizzard and intestine are described as having a variable behaviour towards adrenaline, concentrations above 1:100,000 usually caused a decrease in tone, those below, an increase.

Thus when the effects of similar concentrations of adrenaline are compared in the two sets of experiments, there is a sharp difference, for whereas in the majority of cases described by Wu, 1:1,000,000 adrenaline causes an increase in tone of the crop gizzard and intestine, in the experiments on the same regions of the gut described here (with the sole exception of the four described on p. 363), 1:1,000,000

adrenaline causes a fall in tone. This difference may possibly be related to the different technique employed in the two sets of experiments, for Wu did not record the changes in tone of portions of the intact gut, but of an isolated piece, suspended vertically in a bath of special design (Wells 1937*a*). Consequently in Wu's experiments the movements directly recorded were those of the longitudinal muscles, whereas in the experiments described here the movements directly recorded were those of the circular gut muscles. These results therefore may possibly indicate reciprocal innervation of the circular and longitudinal muscles. Again, in Wu's experiments, the drug approached its seat of action by diffusion from the surrounding fluid in the bath, but in the experiments described here the drug approached its seat of action via the vascular system.

The effect of ephedrine resembles that of adrenaline (figures 7, 8). This resemblance is also described by Wu in the case of the buccal cavity and pharynx, and crop and gizzard, which he found to react to ephedrine (1:10,000–1:100,000) in the same way as to the stronger solutions of adrenaline (down to 1:100,000). Several workers on vertebrates (see Gaddum 1938), in addition to describing similarities in the action of the two drugs, also found that low concentrations of ephedrine potentiate the action of adrenaline. Wu also found this to be so in the earthworm, and suggests that the action of ephedrine by itself may possibly be due to its potentiating effect on the action of an adrenaline-like substance already present (p. 190), comparing this action of ephedrine with that of eserine when administered alone. However, it must be noted that the potentiating effect of eserine is held to depend on its inhibiting action on an enzyme which destroys acetylcholine (Dale 1914; Englehart & Loewi 1930; Matthes 1930; Plattner & Hunter 1930; etc.). This enzyme is easily demonstrable in the earthworm (see below), but amine oxidase, which has been suggested as the enzyme responsible for the destruction of adrenaline and related compounds, and which is inhibited by ephedrine, as shown by Blaschko, Richter & Schlossman (1937), and Gaddum (1938), has not been found in the earthworm (Blaschko *et al.* 1937).

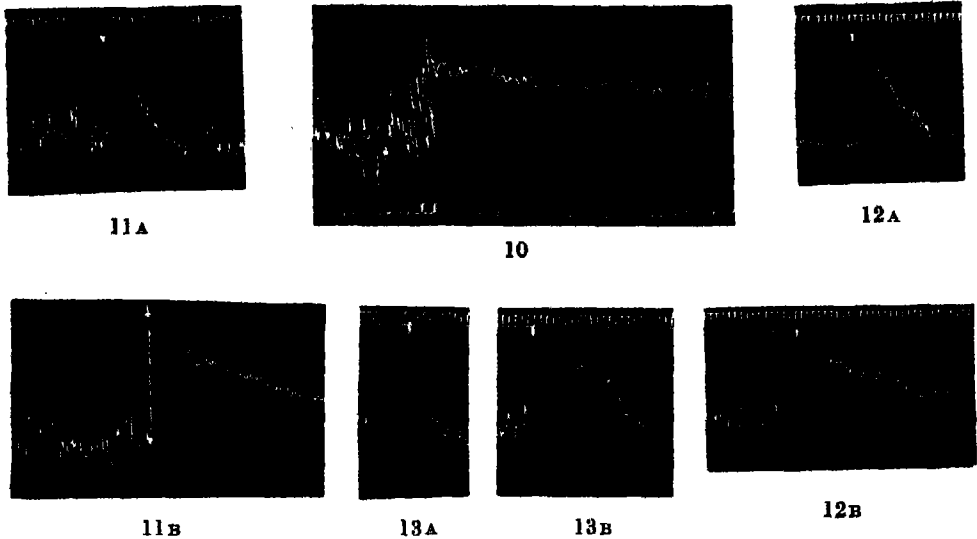
The similar action of tyramine (figure 9) and adrenaline is also in harmony with their action in vertebrates.

THE ACTION OF PARASYMPATHETICOMIMETIC DRUGS—ACETYLCHOLINE

Methods

In the majority of experiments the technique was the same as that employed in the experiments with adrenaline. The solution of acetylcholine was obtained by dissolving acetylcholine chloride or bromide (B.D.H.) in phosphate solution in the usual way. In virtue of its stability, the latter was used as a stock solution, being diluted to a strength of 1:1,000,000 by addition of Ringer immediately before use. When eserine was used, the solution was prepared by dissolving sufficient physostigmine salicylate or physostigmine sulphate in Ringer, to make a 1:10,000 solution. The drugs were injected in doses of 0.02 c.c.

In later experiments it was found expedient to replace the method of injecting acetylcholine after eserine by one in which the acetylcholine was discharged from a pipette directly on to the gut. It had been ascertained previously by means of controls, that administering the drug in this way produced an effect not significantly different from that produced by injection.



FIGURES 10-13. Effect of parasympatheticomimetic drugs on various regions of the gut.
Drug administered at signal.

FIGURE 10. Oesophagus: 1:10,000 physostigmine salicylate; 2nd signal, 1:1,000,000 acetylcholine.

FIGURE 11. A, Oesophagus at signal, 1:1,000,000 acetylcholine; B, repetition of last experiment after 1:10,000 eserine. (Drum speed as in figure 10.)

FIGURE 12. A, gizzard: 1:100,000 acetylcholine; B, repetition of last experiment after injection of 1:10,000 eserine.

FIGURE 13. A, intestine: 1:100,000 acetylcholine; B, repetition of last experiment after previous injection of 1:10,000 eserine.

Acetylcholine produced a rise in tone in all regions of the gut examined (figures 10-13). Control experiments in which the diameter of the oesophagus was measured before and after administration of acetylcholine were made as in the previous experiments with adrenaline. Two results, selected to illustrate the considerable range of variation of the effect, are shown in table 2.

It was found that the injection of eserine was sometimes, but not always, followed by a rise in tone of the gut, but even when not followed by a rise, subsequent injection of acetylcholine produced a rise in tone which was both greater and more prolonged than when the latter was injected alone (compare figures 11A, 11B). Eserine therefore exerts its usual potentiating effect on the action of acetylcholine.

TABLE 2

One of the hearts injected with approximately 0.02 c.c. 1:10,000 eserine. After 5 min. a further injection of 0.02 c.c. of a 1:1,000,000 solution of acetylcholine was given

selected experiment	diameter of oesophagus in arbitrary units	
	before injection	after injection
1	60	35
2	45	40

The increase in tone produced by acetylcholine is also described by Wu (1939*a*), who used a different technique (see p. 365). He also described a graded series of responses, varying with the concentration of the acetylcholine used.

It is noteworthy that the results do not suggest a reciprocal innervation like those described above.

Thus the effect of injecting acetylcholine is the opposite of that produced in most instances by adrenaline, and the antagonistic action of the two at once recalls that of the augmentor and inhibitor nerves previously described (1943). This suggests that these nerves are cholinergic and adrenergic respectively. It must be borne in mind, however, that the administration of drugs by injection leads to their wide distribution through the body, making possible their action on peripheral relays. The stimulating action of acetylcholine on sympathetic ganglia in vertebrates has been shown by Dale (1914) and Feldberg & Gaddum (1934), and therefore, in experiments such as the foregoing, the possibility of the drugs acting on nerve cells which may be situated along the augmentor and inhibitor tracts, as well as on the effectors at the periphery, must always be borne in mind.

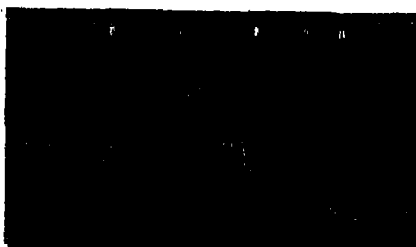
The experiments with acetylcholine and adrenaline have extended the resemblance already shown to exist between the gut of the earthworm and that of the vertebrates (Millott 1943), for the effects of the augmentor and inhibitor nerves, like those of their counterparts in the vertebrates, can be produced by acetylcholine and adrenaline respectively. However, the fact that the effect of the augmentor and inhibitor nerves can be reproduced by acetylcholine and adrenaline is insufficient to warrant the conclusion that these groups of nerves are cholinergic and adrenergic. Accordingly, further evidence was sought.

FURTHER EVIDENCE OF CHOLINERGIC AUGMENTOR NERVES

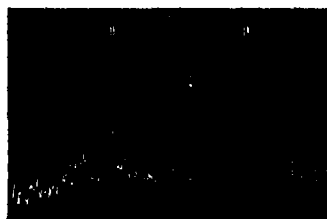
Many of the effects of acetylcholine in vertebrates are known to be abolished by atropine. If, therefore, the effect of stimulating the augmentor nerves is abolished by injection of atropine, it provides strong evidence that these nerves are cholinergic.

It has been shown in a previous communication (Millott 1943) that the nerves which augment the tone of the gut traverse the ventral aspects of the septa. If, therefore, it can be shown that stimulation of these nerves is ineffective after

injection of atropine, then evidence has been obtained which indicates that the augmentor nerves are cholinergic. The experiment shown in figure 14 illustrates this. The augmentor nerves in a septum of the oesophageal region were stimulated by means of fine silver-wire electrodes applied to the ventral aspects of the septum. The electrodes carried a 4 V alternating current drawn from a mains transformer and controlled by a 400 ohm potentiometer. Precautions against short circuiting were taken as described in a previous account (Millott 1943). The pronounced rise in tone of the oesophagus produced by stimuli, such as those given at the first two signals, is abolished after injecting 0.05 c.c. of a 1:1000 solution of atropine (at third signal), as shown by the lack of response to repetition of the stimuli (at 4th and 5th signals).



14



15

FIGURES 14, 15. Evidence of cholinergic augmentor nerves. In both cases electrical stimulation was achieved by the circuit described on p. 369.

Figure 14 illustrates the abolition of the augmentor response of the oesophagus by means of atropine. Signals 1 and 2, electrical stimulation of septum. Signal 3, injection of atropine. Signals 4 and 5, repetition of stimuli. Figure 15 illustrates potentiation of the augmentor response of the oesophagus by eserine. Signal 1, weak electrical stimulation of the peritoneum for 4 seconds. At arrow, 1:100,000 eserine injected into heart. Signal 2, repetition of stimulus.

In the same way, if the effect of stimulating augmentor nerves is potentiated by eserine, there is further good reason to believe that the nerves are cholinergic. The previous communication has shown that the peritoneum lining the body wall of the earthworm is traversed by nerves which augment the tone of the gut. If, therefore, the effect of stimulating the peritoneum on the tone of the gut is increased by injection of eserine, further indication of cholinergic augmentor nerves has been obtained. Figure 15 illustrates this; the rise in tone of the oesophagus produced by stimulating electrically the peritoneum lining the neighbouring body wall for 4 sec. (timed by stopwatch) is greatly enhanced in extent and duration, after the injection of eserine (2nd signal).

Again, if the existence in the earthworm of a substance which rapidly destroys acetylcholine can be shown, the indications of a cholinergic nervous system will be further confirmed. The following experiment affords such confirmation. An extract of the gut wall of a worm was made by grinding about 0.35 g. of the cleaned gut with clean sand and 1.5 c.c. Ringer. The filtered extract was allowed to stand

with a 1:1,000,000 solution of acetylcholine (at pH 8.1) for $\frac{1}{2}$ hr. The fluid was then tested for acetylcholine by a method based on that of Minz (1932) and Chang & Gaddum (1934). A strip of the dorsal body-wall muscle of *Hirudo medicinalis*, previously eserinizd (Fuhner 1918 *a, b*), was suspended in a bath containing 3.5 c.c. Ringer. The Ringer was continually oxygenated by a fine stream of air, and the bath was arranged so that the contained fluids could be readily withdrawn or renewed. Crystals of physostigmine salicylate were added to the bath to make a concentration of about 1:100,000, and to allow the drug to exert its full effect, the preparation was allowed to stand for half an hour. The solution of acetylcholine, which had been mixed with extract of gut wall, was added to the bath. A control was performed in which the gut-wall extract was replaced by an equivalent amount of pure Ringer, and the effect on the same piece of eserinizd leech muscle was determined. The results of the experiment are shown in figure 16. The contraction produced by acetylcholine solution previously mixed with gut-wall extract (figure 16A) was very small compared with that produced by the solution in the control experiment (figure 16B).

Thus a substance is present in the gut wall of earthworm which destroys the activity of acetylcholine.



FIGURE 16. Demonstration of cholinesterase. A, response of eserinizd leech muscle to 1:1,000,000 acetylcholine previously in contact with an extract of gut wall; B, control experiment with same muscle in which gut wall extract was replaced by Ringer.

Finally, the most convincing evidence rests on the fact that it is possible to extract from fresh whole earthworms a substance which possesses pharmacological properties strikingly similar to those of acetylcholine. The extract was prepared by the method of Chang & Gaddum (1934), and tested by eserinizd leech muscle as described above. Extracts were found to contain a highly active substance, 0.3 c.c. producing the powerful contraction seen in figure 17B. The following two tests show the close similarity between the effects of this substance and those of acetylcholine. First, the extract has no effect on leech muscle which has not been previously eserinizd (figure 17A). Acetylcholine has a relatively weak effect on leech muscle untreated with eserine, which greatly increases its sensitivity (Chang & Gaddum). Secondly, the activity of the extract was destroyed by previous mixture with fresh frog's blood, owing to the cholinesterase present in the blood (figure 17C).

Thus the indications of the experiments with acetylcholine, described in the early part of the account, are fully confirmed by the experiments just described, and there is now substantial evidence of cholinergic augmentor nerves passing to the gut via the segmental nerves and septa.

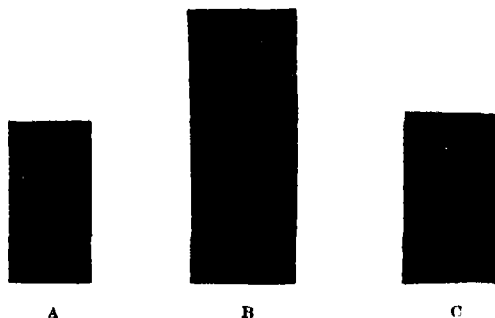


FIGURE 17. Pharmacological properties of a substance extracted from earthworm. A, effect of extract on leech muscle; B, effect of same extract on eserinizied leech muscle; C, effect of same extract, previously mixed with fresh frog's blood, on eserinizied leech muscle.

DISCUSSION AND CONCLUSIONS

The evidence just presented affords a clear indication of chemical transmission of nerve impulses to the alimentary muscles of earthworm.

Evidence of chemical transmission in invertebrates has already been brought forward by several workers: notably Gaskell (1914) in *Hirudo*; Bacq (1934, 1935, 1937) in Turbellarians, Nemertines, Annelids, Sipunculids, Molluscs, Echinoderms and Crustacea; Wells (1937*a, b*) in *Arenicola*; Wu (1939*a, b*) in *Lumbricus*; Welsh (1939) in Crustacea. It is noteworthy, however, that where evidence of cholinergic transmission has been described, it does not necessarily exhibit all the characters associated with this type of transmission in vertebrates. Thus Welsh (1939) has brought forward evidence of cholinergic transmission in Crustacea in which there is little or no cholinesterase in the blood (Simonart 1931; Bacq 1937). The close similarity between the manifestations of chemical transmission in earthworm and vertebrates is therefore noteworthy.

In instances where there is evidence of adrenergic transmission in invertebrates, it has sometimes been associated with a specialized activity or system of the animal. Thus Gaskell (1914) showed that nerves which show signs of being adrenergic, play a vital part in controlling the specialized system of contractile coelomic channels ('blood vessels') in *Hirudo*, and Wells (1937) has shown that adrenergic nerves have a special significance in initiating the continuous oesophageal rhythm associated with the characteristic flight reaction of *Arenicola*. Chemical transmission in earthworm has, as yet, shown no such special significance.

The most significant outcome of the foregoing investigation, however, is that it makes possible an even closer comparison between the systems of indirect

control of alimentary tone in earthworm and vertebrates. The previous investigation (Millott 1943) has already shown that the gut of the earthworm, like that of a vertebrate, is indirectly controlled by two antagonistic groups of nerves, and there is now a clear indication that the two groups of nerves in earthworm, as in the vertebrate, are adrenergic and cholinergic.

It is very interesting to note that Wu (1939a) also comes to the conclusion that the gut of the earthworm is influenced by a system of antagonistic adrenergic and cholinergic nerves, but he restricts this type of innervation to the posterior portion of the gut, whereas in the previous account given by me (1943), this double innervation is shown to influence all regions of the gut behind the pharynx. From the description given by Wu of his experiments, it is clear that they are based on a conception of a segmentally arranged nerve supply to the gut which does not involve nerves in the body wall, previously shown by me to play a prominent part in the indirect control of the gut. He believes that the nerves arising from the ventral nerve cord reach the gut by a more direct route, viz. via the segmental nerves and septa and segmental blood vessels. No experiments are described in which stimulation of the segmental nerves individually, or the body wall, was undertaken, and consequently Wu was not able to demonstrate the separate action of the inhibitor and excitor nerves arising from the central nervous system, nor the separate action of the somatic nerve plexuses, described in my previous paper. He thus remains uncertain of the existence of inhibitor nerves to the pharynx and oesophagus, and since he does not obtain inhibition of these portions of the gut with acetylcholine or adrenaline, he concludes that if inhibitors do exist, their impulses are transmitted to the effectors by some other substance.

The foregoing experiments and the previous account show that such a conclusion is unwarranted, and that the nerves controlling the oesophageal tone are essentially similar in arrangement and function to those supplying the more posterior regions.

Despite the clearness of the resemblances between the type of indirect control of the gut in earthworm and vertebrates, it must be borne in mind that they exist as broad generalizations only, and that there are the many marked differences described in a previous paper (Millott 1943). Much more experimental work on the control of other autonomic effectors must be performed before the similarities discovered can be given their true significance. It is hoped to undertake such work in the future.

I wish to thank Dr C. F. A. Pantin, F.R.S. for his invaluable and constant help, and Professor H. Graham Cannon, F.R.S. for his continued interest and advice. My thanks are also due to Professor J. Gray, F.R.S. and Professor J. Stanley Gardiner, F.R.S., in whose Department a large part of the work was carried out. My thanks are expressed to Dr R. Blaschko for many helpful suggestions, and to Dr D. Richter. The greater part of the work was carried out whilst holding a Rouse Ball Research Studentship of Trinity College, Cambridge.

REFERENCES

- Bacq, Z. M. 1934 *Arch. int. Physiol.* 38, 138.
- Bacq, Z. M. 1935 *Arch. int. Physiol.* 42, 24.
- Bacq, Z. M. 1937 *Arch. int. Physiol.* 44, 174.
- Blaschko, H., Richter, D. & Schlossman, H. 1937 *Biochem. J.* 31, 2187.
- Chang, H. C. & Gaddum, J. H. 1934 *J. Physiol.* 79, 255.
- Dale, H. H. 1914 *J. Pharmacol.* 6, 147.
- Englehart, E. & Loewi, O. 1930 *Arch. exp. Path. Pharmac.* 150, 1.
- Feldberg, W. & Gaddum, J. H. 1934 *J. Physiol.* 80, 12P.
- Fuhner, H. 1918a *Arch. exp. Path. Pharmac.* 82, 44, 81.
- Fuhner, H. 1918b *Biochem. Z.* 92, 347.
- Gaddum, J. H. 1938 *Brit. Med. J.* 1, 713.
- Gaskell, J. F. 1914 *Phil. Trans. B*, 205, 153.
- Matthes, K. 1930 *J. Physiol.* 70, 338.
- Millott, N. 1943 *Proc. Roy. Soc. B*, 131, 271.
- Minz, B. 1932 *Arch. exp. Path. Pharmac.* 91, 342.
- Plattner, F. & Hunter, H. 1930 *Pflug. Arch. ges. Physiol.* 225, 19.
- Simonart, A. 1931 *Rev. belge. Sci. méd.* 3, 1.
- Smallwood, W. M. 1923 *Proc. Nat. Acad. Sci., Wash.*, 9, 99.
- Wells, G. P. 1937a *J. Exp. Biol.* 14, 117.
- Wells, G. P. 1937b *J. Exp. Biol.* 14, 290.
- Welsh, J. H. 1939 *J. Exp. Biol.* 16, 198.
- Wu, K. S. 1939a *J. Exp. Biol.* 16, 184.
- Wu, K. S. 1939b *J. Exp. Biol.* 16, 251.

The origin of the primitive tetrapod limb

BY T. S. WESTOLL, PH.D., D.Sc., F.G.S.

Department of Geology and Mineralogy, University of Aberdeen

(Communicated by D. M. S. Watson, F.R.S. Received 18 February 1943.)

A rhipidistian paired fin, made to assume the position of a primitive tetrapod limb, would suffer certain modifications which throw light on the origin of the latter and on the homology of its elements. The new theory here stated, and that of Gregory and Raven (from which it differs in several points), are radically different from all hitherto proposed.

The proto-tetrapod pectoral and pelvic fins were of similar structure. Each consisted of five axial mesomeres, each bearing distally a simple preaxial radial and an actual or potential postaxial process; a rudimentary sixth mesomere may have been present.

The downward flexure of the extended fin between 1st-2nd mesomeres, and forward torsion of the part distal to the 2nd mesomere, leads to conditions closely comparable with part of the limbs of primitive tetrapods. The digits (including their proximal podialia) are new formations, and the terms archepodium and neopodium are proposed for those parts of the tetrapod limb derived directly from the rhipidistian paddle skeleton and for new formations respectively. The prepollex (prehallux) is a digit, but postminimal and pisiform elements are less certainly so. The main axis of the rhipidistian paddle is directed between the 'first' and 'second' fingers. Several variants of this scheme are discussed.

The structure of the stegocephalian limb is discussed; it corresponds well to this theoretical pattern, but there is some reduction of the distal part of the archepodium.

The limb structure of urodeles and other tetrapods can easily be derived from this common ancestral pattern, contrary to the views of Holmgren. The development of limbs of living

tetrapods is in general agreement with the main results of the new theory, which alone explains the relationships of the radiale (tibiale) in many forms.

Many peculiarities of the tetrapod limb seem to be relics of characters of the rhipidistian fin.

Each mesomere and preaxial radial of the rhipidistian paddle was probably related to one myotome. Relics of this condition seem to be found in the segmental motor innervation of the muscles of human extremities.

The relationship between paired fins in Dipnoi, Coelacanthini and Rhipidistia is discussed.

I. INTRODUCTION

The long-standing problem of the evolution of the 'cheiropterygium' from a fish fin is now much nearer solution with the positive recognition of the fish ancestors of the tetrapods among the rhipidistian Crossopterygii (cf. Westoll 1943 for review). The structure of the paired fins in these forms is known in *Sauripterus*, *Tristichopterus* and *Eusthenopteron*, and less completely in *Megalichthys*, *Ectosteorhachis* and a few other forms, which seem to follow the same general pattern. The earlier comparisons with sharks and with *Polypterus* (which is certainly an actinopterygian) need no longer be quoted. The general similarity between the skeleton of the rhipidistian paddle and the tetrapod limb has been recognized since the time of Hall; nevertheless, the exact correspondence of parts which might have been expected is not immediately obvious, and there have been many different attempts to relate the two types of limb structure—e.g. by Watson (1913), Broom (1913), Gregory (1915), Petronievics (1918), Gregory, Miner & Noble (1923), Romer & Byrne (1931), Holmgren (1933, 1939), Gregory (1935), Howell (1935, etc.), Steiner (1935), and others.

The earlier accounts are essentially concerned with the discovery by formal comparison of the 'main axis' of the supposedly primitive archipterygium in the tetrapod limb, and of homologues of the usual five tetrapod digits in the fish fin; each finger was usually homologized with a single endoskeletal ray. Most authors who have used the rhipidistian comparison agree in finding the main axis near the postaxial side of the tetrapod limb, e.g. between digits III–IV (Steiner), in the line of digit IV (Watson, Gregory, 1935), between IV and V (Petronievics), along V (Gregory *et al.* 1923), or in the pisiform (Broom). Others, starting from a theoretical or dipnoan biserial archipterygium, have found the main axis in digit III (Huxley, Marsh), or II (Goette, Wiedersheim), while Holmgren finds it between I and II in urodeles. Romer & Byrne (1931) considered that the previously accepted orientation of the rhipidistian paddle is incorrect, and on their scheme the element usually called 'radius' in *Eusthenopteron* and *Sauripterus* is the ulna, and the main axis must be near the preaxial margin. Howell (1935) follows the same plan of orientation, and finds the axis of the rhipidistian limb in the first finger of the tetrapod, though he mistakenly places the radiale and precentrale (see later) of *Eryops* as parts of digit I. Steiner (1935, p. 717) sums up: 'Kurzum, es ist nicht eine Konstruktion unversucht geblieben, um die Tetrapodenextremität auf die verschiedenartigen Radiensystemen eines uni- oder biserialen Archipterygium zu beziehen'.

Holmgren seems to have been the first to insist (1933) on the importance of the fin-like early stages of the developing limb of some urodeles, and noted that the tip of the fin, and thus presumably the main axis, lies in these forms between digits I and II.

Gregory & Raven have recently (1941) suggested a scheme more satisfactory than any so far proposed, in which they find the metapterygial axis running through humerus, intermedium and two centralia to emerge between digits I and II. I had come to somewhat similar conclusions from analysis along the lines of the present paper, and gratefully acknowledge the stimulus of Gregory & Raven's account. However, those authors do not seem fully to have appreciated the mechanism of the transition, and their theory in detail is not in complete accord with the evidence from embryology of living forms.

2. TRANSFORMATION OF RHIPIDISTIAN PADDLE INTO WALKING LIMB

I have for some time been investigating the paired fins of early fishes, and have given much attention to the Crossopterygii and Dipnoi. It has always seemed to me that the process of transformation to a walking limb involved fundamentally only the differential emphasis of movements and of structures already characteristic of, or possible in, the fish fin, and that discontinuous changes (such, for example, as those postulated by Romer & Byrne (1931)) are most unlikely. One of the first difficulties has been to understand the orientation and range of movement of the rhipidistian paddle. Briefly, it is now clear that the pectoral and pelvic fins were similarly held, that the preaxial margin of each fin is usually marked by stronger dermal rays, and that the main axis of mesomeres is not preaxial. The fin is often preserved applied to the flank, preaxial margin downwards, flexor surface turned to the flank; in this position the plane of the fin must have been nearly vertical (figure 1A, B). From this 'trailing' position the fin could be extended laterally, when it was probably normally held nearly in the horizontal plane (figure 1A). This movement involves therefore a rotation of the plane of the fin about its long axis, a movement comparable in kind with, and probably rather similar in extent to, the extension of a bird's wing. In dorso-ventral compressions seen from the ventral side, the paired fins may be preserved in two main orientations—if in the trailing position, the preaxial margins are mesial, but if extended, the preaxial margins are antero-lateral. It appears therefore that the pectoral fin, when applied to the flank in the trailing position, had its preaxial edge downward, a condition rare in living fishes; *Ceratodus*, *Polypterus*, *Amia* and other forms all have the preaxial edge upwardly directed when the fin is trailing. On the other hand, the pelvic fins of several of these forms resemble those of Rhipidistia in their orientation when trailing. The rhipidistian pectoral and pelvic fins, then, were held in a similar manner; my observations on *Scaumenacia* and other early Dipnoi suggest that the same was true for them.

The mode of use of such fins is somewhat uncertain in view of the absence of full

information concerning the musculature; I have not so far been able to make satisfactory preparations of uncrushed material of either the scapula or the pelvic girdle. In view of the narrowness of the proximal part of the fin skeleton it is probable that considerable rotation about the fin axis could be induced, and a rather wide range of functions was no doubt possible. It is almost certain that the fins were used both as keels (when fully extended in the horizontal plane) and as brakes (when partly extended in a vertical plane). They might also perhaps have been used for slow paddling locomotion, the forward (recovery) stroke involving rotation to the horizontal plane and extension, the backward (propulsive) stroke involving rotation to the vertical plane and retraction. If the fin were sufficiently long the type of movement just mentioned could easily become a rough and ready means of contact locomotion, but it would be very inefficient since the force would bear on the delicate edge of the fin. Any attempt at efficient 'crawling' would appear to involve a different technique; a more extensive forward swing (or a more forward placing of the bearing region of the limb at the beginning of the propulsive stroke) and a larger bearing surface would be almost essential; and any slipping, sliding or skidding affecting the bearing surface would be very wasteful of energy. It is clear, from the study of fossil footprints of early tetrapods, that the bearing surface did not suffer excessive skidding or rotational slipping during the stride, and it will be as well to consider the attainment of this condition from the start.

There are two main types of endoskeleton of the rhipidistian paddle. The pattern of *Eusthenopteron* is repeated in its close ally *Tristichopterus*, and apparently also in *Megalichthys*; in it each preaxial ray is simple. The *Sauripterus* fin is somewhat more elaborate, and the preaxial rays show more or less dichotomous branching. Many investigators have chosen the *Eusthenopteron* pattern as a basis for argument, others have used *Sauripterus*. For reasons which will be clear later on, the simpler type is taken here as the basic pattern; the *Sauripterus* paddle is very broad, and it is reasonable to suppose that the type ancestral to the tetrapod limb was not so expanded. In *Eusthenopteron* there are five 'axial' mesomeres in the pectoral, four in the pelvic fins; in the *Sauripterus* pectoral there are four (or five). A fin suitable for transformation into a walking limb would probably be rather elongated; as will shortly be clear, the initial assumption will be made that the ancestral pectoral fin had six mesomeres (five well developed and one rudimentary); and since the tetrapod hand and foot are constructed on essentially similar lines, the same structure will be taken as basic for the pelvis.

Such a fin is shown diagrammatically, abducted from the body, in figure 1A, r. Each mesomere is drawn out distally and postaxially into a short process, and bears a preaxial radial, besides articulating with a more distal mesomere (if present). The postaxial process is not present on every mesomere of *Eusthenopteron*; it is here regarded as potential even where not developed. Moreover, in *Ceratodus* (and probably in some Rhipidistia, e.g. in *Sauripterus*) separate postaxial rays develop in the same position as these processes on some mesomeres; this is likewise regarded as a potential development in Rhipidistia, since the Rhipidistia and Dipnoi had

probably not long separated from a common ancestor in the Middle Devonian. Such a fin structure may now be considered taking up a position affording a broad bearing surface, that is, with the proximal segment directed laterally, the more distal part bent down (figure 1 c); the flexor surface thus splays out on the 'ground'.

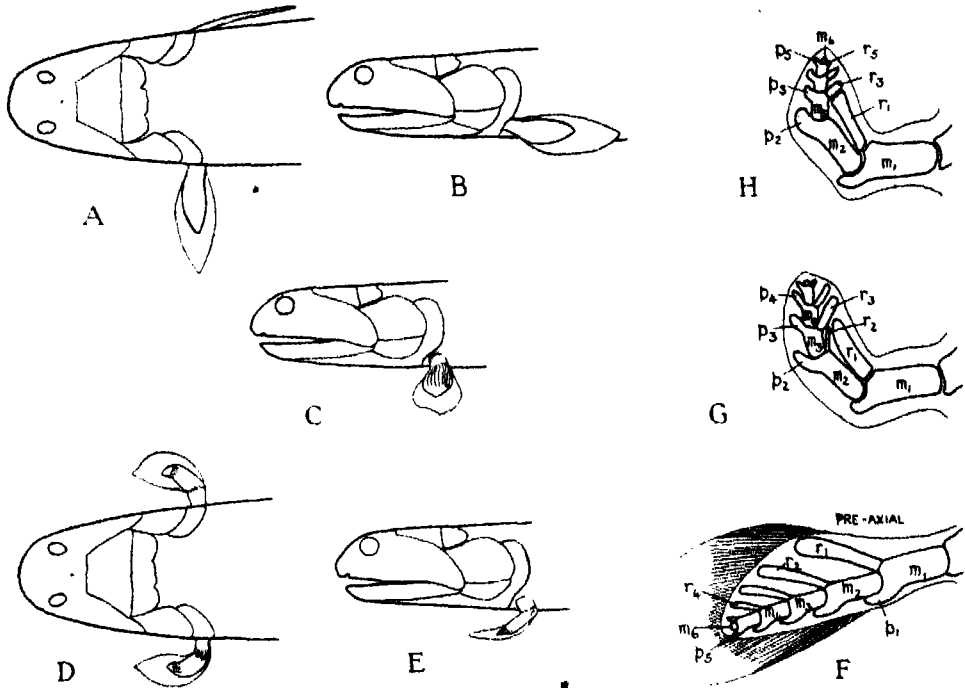


FIGURE 1. Possible changes in the rhipidistian paddle during development of a walking habit. A, dorsal view; right paddle trailing, left paddle fully abducted in horizontal plane. B, lateral view; left paddle in trailing position. C, lateral view; left paddle bent down at 'elbow', flexor surface at contact with bearing surface. D, E, dorsal and lateral views; paddles rotated into proto-tetrapod positions. F-H, three possible stages in structural modification of the skeleton of the paddle.

Explanation of lettering for figures 1-4: *c.d.*, *c.p.* distal and proximal axial centralia; *F.* fibula; *fe.* fibulare; *H.* humerus; *i.* intermedium; *m₁₋₆*, mesomeres; *p₁₋₅*, postaxial processes of mesomeres; *pod.*, *pcp.* distal and proximal post-centralia; *Ph.* prehallux; *pis.* pisi-form; *pm.* postminimus; *Pp.* prepollex; *ppm.* post-postminimus; *pr.c.* precentrale; *R.* radius; *r₁₋₅*, preaxial rays of mesomeres; *rle.* radiale; *T.* tibia; *te.* tibiale; *U.* ulna; *Ule.* ulnare; *x.* proliferations of radius anlage; I-V, digits of the 'pentadactyl limb'; 0, carpal of prepollex (tarsal of prehallux); 1-5, carpalia of digits.

This would allow propulsion, though only with a slithering motion; but a forward turning of the distal part would allow a fixed tread and a free stride. This inturning would make it almost impossible for the second preaxial ray to develop at all because of interference from the first, and would bring the more distal preaxial rays across the line of the first preaxial ray (figure 1 G, H). The loss of the second preaxial ray, and the inturning of the main axis in front of the first preaxial ray, were deduced also by Gregory & Raven (1941, figure 25), but the writer diverges

Two of these possible patterns are specially examined in this paper. According to one (figure 2A) the archepodium includes (i) separated postaxial processes, represented by the ulnare (fibulare) and both postcentralia (for definition see below), and (ii) preaxial rays, now radius (tibia), the radiale (tibiale), and the precentrale (for definition see below), and the digits alternate in position with the distal elements of the archepodium. The other discussed possibility (figure 2A') is that the digits were originally linearly related to the more proximal elements; the radiale (tibiale) and precentrale are shown in the same axis of development as the radius, and may not therefore be derived from preaxial rays and may belong to the neopodium. This raises the suspicion that the postcentralia and ulnare (fibulare) may be similarly related to digits III-V, and may also be neopodial structures. For various reasons this does not seem to be very probable. A third pattern (figure 2A'') is a compromise in which the digits are regarded as linearly related to elements of an archepodium essentially similar to that of figure 2A. Another possible pattern, not illustrated, is similar to that shown in figure 2A, but with the radiale and precentrale considered as lying in the line of the radius as in figure 2A'', and therefore possibly neopodial structures.

In the reconstruction at present favoured by the writer (figure 2A) the postaxial process of the humerus (femur) does not separate, but may become an important muscle attachment (entepicondyle of humerus); the similar processes of the more distal mesomeres separate as the ulnare (fibulare), proximal postcentrale and distal postcentrale. The preaxial ray of the first mesomere becomes the radius (tibia), that of the second is suppressed (cf. figure 1F-H). That of the third becomes the radiale (tibiale), and of the fourth the precentrale, though there are difficulties to be met in this interpretation. The new terms postcentrale and precentrale are here introduced for the elements currently known in living tetrapods and in *Stegocephalia* as follows (discussion later):

- | | |
|------------------------|--|
| Precentrale | = element Y (centrale praepollicis or praehallucis) of Steiner (1921) and Holmgren (1933, etc.). |
| | = centrale 1 of the stegocephalian carpus (Gregory & Raven 1941) and tarsus (Schaeffer 1941). |
| Postcentrale proximale | = centrale ulnare (fibulare) proximale (Steiner 1921; Holmgren 1933, urodeles only). |
| | = centrale 4 of other tetrapods (Holmgren 1933). |
| | = (not separately present in known <i>Stegocephalia</i>). |
| Postcentrale distale | = centrale ulnare (fibulare) distale (Steiner 1921; Holmgren 1933, urodeles only). |
| | = centrale 3 of other tetrapods (Holmgren 1933). |
| | = centrale 3 of the stegocephalian carpus (Gregory & Raven 1941) and tarsus (Schaeffer 1941). |

The remaining centralia (remains of the original axis) are here called simply centrale proximale and centrale distale, more strictly 'centrale axiale' in each case;

they are termed centralia radialis (tibialis) by Steiner and others, and are respectively the 'centrale 1' and 'centrale 2' of Holmgren's notation in non-urodele tetrapods and 'centrale 4' and 'centrale 2' of Gregory & Raven (1941) and Schaeffer (1941).

This theory is the only one that accounts for the fact that the radiale (tibiale) and precentrale often seem to originate in ontogeny, not as distal proliferations from the radius-anlage, but as side-branches from the main axis (cf. figures of Holmgren 1933, Steiner 1934, 1935 (figure 3f), and figure 4g, j of this paper). On the other hand, most urodeles seem to show these elements aligned with the radius even at an early stage of development, which may prove to be evidence against this part of the theory, and would agree better with the conditions in this region shown in figure 2A''. However, there is reason to believe that in several vertebrate structures in which great changes in form and function can be proved (e.g. in the history of the mammalian jaw and ear-ossicles), organs or structures may come to lie in the embryonic field of influence of other organs or structures so that an intimate linear developmental connexion may occur, even where the ancestral organs or structures were quite independent. In this specific case there would be good mechanical reasons for a readjustment of elements in the final fashioning of the walking limb. Alternatively, of course, the shortening of the carpus and tarsus in higher tetrapods may have been accompanied by a rearrangement in which the radiale and precentrale became more closely connected with the old main axis.

Little material has been available to the writer for the first-hand examination of early stages of limb development in living tetrapods, but the figures of Holmgren (1933, 1939), with those to which he refers, afford a fairly comprehensive atlas. Comparison with the stegocephalian limb is now more certain, since the new descriptions of the forelimb of *Eryops* (Gregory & Raven 1941) and the hindlimb of *Trematops* and of *Archegosaurus* (Schaeffer 1941). Comment on certain points raised by these works is incorporated in the following discussion.

(a) *The relations of the digits to the archepodium*

Here the 'digit' is taken to include phalanges, metapodial and podial (carpal, tarsal). There is good evidence from embryology that the digit so defined is a single recognizable linear unit. In young stages of many tetrapods the blastematos rudiments of the digits are markedly separate from the rest of the developing limb skeleton, and one frequently gets the impression that the digital blastemas are more thoroughly separate on the radial (tibial) than on the ulnar (fibular) side. The outer (ulnar or fibular) side of the hand or foot sometimes shows the digital blastema apparently 'shading off' proximally, lateral to the ulna or fibula. The relationship of the podials to the more proximal limb skeleton is interesting. Holmgren (1933, figures 37, 50, 51, 56) shows, in diagrammatic form, his view that the 'centralia' are generally speaking in line proximally with the digits, though sometimes two digits may converge on a single central (cf. also figures 2A', A'' of this paper). My own impression, derived from examination of actual material and of published

figures of developmental stages of living forms, and of the stegocephalian limb, is that the carpalia or tarsalia primarily alternated with the proximal elements, keying between them. These factors find a reasonable interpretation if the digits are new formations in the tetrapods, primarily conditioned by the arrangement of the skeletal elements of the rhipidistian paddle after its modification to form the archepodium. This archepodium would bear on the flexible region where the muscular lobe of the paddle passed into the web, as pointed out by Gregory & Raven (1941).

The mesoderm in which the digital blastemas form may have had little to do with the mesoderm of the main fin structure—there is evidence that it is postaxial mesoderm. Figure 2A indicates the initial position of the possible digital rudiments on this basis; it will be observed that they are shown alternating with the postaxial processes (ulnare or fibulare, and the postcentralia), the distal axial centrale, and the precentrale; no element is shown between the precentrale and the radiale (tibiale) because of the mechanical conditions of this region, in which these elements are brought round in front of the radius or tibia. The prepollex (or prehallux) falls readily in series as a digit on this scheme. The fifth mesomere is shown with rudimentary postaxial and preaxial rays or processes, and the rudiment of a sixth mesomere (cf. the conditions of the 3rd and 4th mesomeres of the pelvic fin of *Eusthenopteron*—Goodrich 1901; Gregory & Raven 1941, figure 33), and all the known digits can readily be accommodated. The question of the postminimals and the pisiform will be discussed later.

On the alternative scheme outlined in figure 2A'', the digits are in series with the more proximal elements, and the fifth mesomere needs merely to show incipient bifidity at the tip. The radiale (tibiale) and precentrale may be, on this scheme, new formations in tetrapods, with perhaps also the ulnare (fibulare) and postcentralia.

The embryological evidence of the distinctness of the digital blastemata suggests that the elements just mentioned are more intimately connected with the main axis. If, therefore, they are not part of the archepodium they should represent secondary proliferations of archepodial structures, distinct from the digits; the radiale and precentrale would then be related to the radius-anlage.

Such a view is certainly tenable, but it is less 'economical' and therefore (to my mind) less likely, and fails to explain the remarkable frequency of close developmental connexion between the radiale (tibiale) and precentrale and the intermedium and axial centralia.

Several other possible schemes have been tried (e.g. with the radiale or tibiale and precentrale considered as preaxial rays associated with the 4th and 5th instead of the 3rd and 4th mesomeres), but these are for various reasons less satisfactory.

I therefore conclude, with Gregory & Raven, that the digital rudiments were initiated during the evolution of a walking limb, and are not derived from any part of the rhipidistian fin skeleton; but diverge from those authors on the constitution and origin of the region between the radius-ulna (or tibia-fibula) and the digits. This is essentially a difference of opinion concerning the extent of the neopodium.

The nature of the prepollex (prehallux) and pisiform and postminimals is much discussed. Howell (1935), for example, denies that these elements represent digits in ancestral tetrapods; but the prepollex (prehallux) is in full agreement with other digits in its general relationships, and may sometimes consist of two elements—one in the same row as podials 1–5, and a terminal element. It is best regarded as produced under developmental influences similar to those responsible for digits I–V of the usual nomenclature, but always weakly developed. The nature of the 'postminimals' is less clear. In the feet of certain tetrapods (e.g. many urodeles, figure 4E, F), there is a postminimal (*sensu stricto*) lying immediately lateral to the 5th tarsal, plausibly to be regarded as a 'sixth' tarsal; there may also be a post-postminimum which is closely related to the disto-lateral corner of the fibula. The latter may well be of no phylogenetic significance, a possible sesamoid. In the hand only one element is identified—the pisiform. In the ontogeny of some tetrapods (cf. Holmgren 1933 for figures) the pisiform appears to develop lateral to the ulnare, so that it is almost in line with the postaxial carpalia (e.g. in *Telmobatus*), whereas in most forms the pisiform is essentially wedged between the ulna and ulnare. It may therefore correspond in position either to the postminimum or to the post-postminimum of the foot, and may be a secondary (sesamoid) development. There is also the possibility that it represents a true serial homologue of the carpal series in primitive forms (cf. perhaps a postminimum) which has become displaced and specialised in relation to muscular actions.

To sum up, the prepollex (prehallux), though often represented only by a 'pre-carpale' or 'pretarsale', is truly comparable with digits 'I–V', though never so highly developed; but the nature of the various postminimal elements is much more doubtful, and only the true postminimum of the foot is really comparable with the tarsalia. It seems safest to regard the post-postminimum and pisiform as secondary developments, at least in their usual position and relations. On the basis of the scheme indicated in figure 2A, the prepollex (prehallux) and postminimum can easily be understood as serial homologues of the digits, or at least of the podials, but it is noteworthy that the postminimum has no bar to proximal migration. The alternative schemes (cf. figure 2, A', A'') also allow easy understanding of the prepollex as a digital homologue, but there is no place for a postminimum considered as a podial.

In the preceding paragraphs the relationship of the digits to the archepodium has been considered with special reference to the nascent walking limb. The placing and number of the digits seem, at this critical period, to have been determined by the structural and mechanical conditions in the limb imposed by the archepodial skeleton. But once a new region of skeletal differentiation had been produced (neopodium), it is probable that the archepodium (gradually removed from the critical bearing zone of the limb) would no longer exercise so complete a determinative effect on the number and arrangement of digits. In other words, the ontogenetic induction mechanisms determining the digital formula in more advanced tetrapods might come to have but little relationship to the biodynamic conditions

determining the appearance of new skeletal structures in response to the first use of the fish paddle as a walking limb. The effect of reduction of the archepodium on the digits will be shown to favour this view.

(b) *The stegocephalian limb structure*

The most complete material so far described is mentioned above. The analysis of the *Eryops* forelimb by Gregory & Raven (1941, figure 27) seems to be open to some doubt (cf. figure 3A), particularly on comparison with the excellent specimen of *Trematops* figured by Schaeffer (1941, figure 1 and figure 3B of this paper). The position of the pisiform (usual in many tetrapods) shown by them is comparable with that of the post-postminimus, not of the postminimus. Comparison with

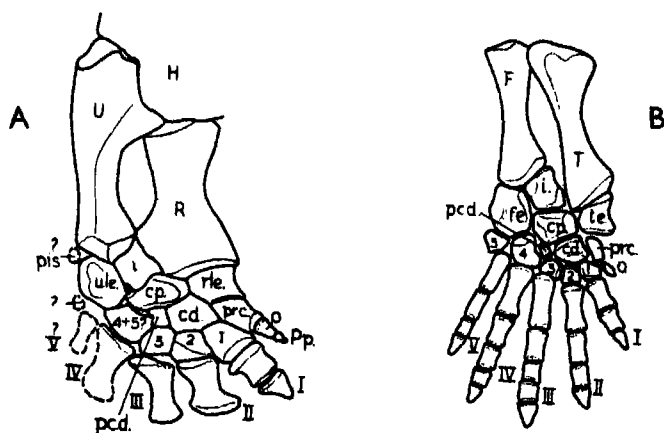


FIGURE 3. Hand and foot in primitive tetrapods. A, flexor view of manus and antebrachium of *Eryops* (slightly modified, by replacement of dislocated digits, from Gregory & Raven 1941). B, extensor surface of pes and leg of *Trematops* (after Schaeffer 1941). (For explanation see legend of figure 1.)

Trematops shows that the carpale labelled '5 + 4' corresponds so well to the 4th tarsale of *Trematops*, that a separate 5th carpale may well have been present if (as Gregory & Raven seem to suggest) there was a 5th digit. It is noteworthy that no living amphibian has a 5th digit in the hand, although *Diplovertebron* among fossil forms preserves it. Schaeffer's account of the *Trematops* foot is most illuminating; it seems fairly clear that, if a postminimus was developed, it articulated with the fibulare adjacent to tarsale 5. His figure and account of *Archegosaurus* (1941, p. 400, figure 2) show the same pattern, though it seems clear that he has misinterpreted part of the structure. He suggests that *Archegosaurus* 'presumably lacks centrale three', but his figure shows clearly that 'centrale 3' (i.e. postcentrale distale) is present, it is tarsale 5 that is missing (cf. the forelimb of *Eryops*!); the fragments regarded as belonging to the first digit are surely incorrectly identified. It is very highly probable, therefore, that the carpus and tarsus of these rhachitomous Stegocephalia were built on a similar plan. With two exceptions—the absence of a separate proximal postcentrale, and the loss of the distal axial element beyond

the corpus of the 5th mesomere or distal axial centrale—this pattern corresponds closely to that deduced above (figures 2B, 3). The position of the postcentrale proximale is occupied by the ulnare (fibulare), carpale (tarsale) 4 (or 4 + 5) and the centrale proximale, all of which are rather large. It is uncertain whether its rudiment has fused with one of these elements or whether it has been lost; it is quite possible that it failed to differentiate from the main axis (cf. conditions in 'normal' and 'abnormal' *Salamandrella*, figure 4E, F). The reduction in the tip of the original axis is continued in other tetrapods.

(c) *The basic pattern of the tetrapod carpus and tarsus*

Holmgren (1933, 1939) insists that the urodele limb shows a structure essentially different from that of other living tetrapods. He derives the former from ancestors with a biserial archipterygium (cf. Dipnoi), the latter from ancestors with a dichotomously arranged fin skeleton which he compares with *Eusthenopteron* and *Sauripterus*; and he attempts to prove that the urodele pattern is not comparable with that of Stegocephalia, contrary to the usual views clearly stated, for example, by Steiner (1921, 1934, 1935). The entire basis of Holmgren's argument is removed by the recognition of the inturning of the mesomeral axis during the evolution of a rhipidistian paddle into a tetrapod limb, and the urodeles actually retain, better than most other living tetrapods, primitive characters shown also in the stegocephalian limbs. Many urodeles show important modifications:

(i) The main axis may be further shortened by loss of the centrale distale, as in *Cryptobranchus* and *Megalobatrachus*, or by even greater reduction (e.g. *Necturus*, *Amphiuma*, *Proetus*).

(ii) There is a marked tendency for reduction of the postcentralia, which often occur only as abnormalities (discussed below).

It is also rather characteristic of urodeles that the radiale and precentrale appear to be parts of the radial ray; this point needs further detailed work before its significance can be assessed (above, pp. 380, 381, and cf. figure 2A''). With regard to the first two points, it is very noteworthy that, except in extreme cases, the reduction in the axial and postaxial centrals is not accompanied by significant modification of the digits. This is doubly important, since it confirms the independent nature of the digital anlagen, and also shows that the absence of traces of the sixth mesomere and distal processes of the 5th in known tetrapods is not a bar to the acceptance as primitive of conditions shown in figure 2A.

The distribution of 'abnormal' elements in the carpus and tarsus of urodeles can be worked out from the examples collected by Holmgren (1939):

(a) In those urodeles with normally only one axial centrale the commonest abnormality is a separate centrale distale. There is not necessarily any disturbance of the digital formula in consequence.

(b) Additional 'centralia' (*s.l.*), if present, occur on the postaxial side of the axial row (intermedium—centrale axiale proximale—centrale axiale distale) and are postcentralia.

(c) These postcentralia, wherever their development is known, are intimately connected with the axial series.

(d) Out of nineteen figured examples with accessory centralia, a post-centrale lies proximal to podialia 3-4 in fourteen cases, proximal to 4 or 4-5 in seven cases; the former is the theoretical position of a distal, the latter of a proximal, post-centrale.

(e) The proximal postcentrale typically meets ulnare (fibulare), intermedium and centrale proximale; the distal postcentrale typically meets the proximal post-centrale (if developed, otherwise the ulnare or fibulare) and the two axial centralia.

Each postcentrale in, for example, *Salamandrella* (figure 4E, F) may be either a true outgrowth of the corresponding axial centrale, or borne, with that element, by the next proximal element of the axis, as the ulna bears ulnare and intermedium. Schmalhausen's plates (1910) show more evidence for the latter possibility than do Holmgren's line drawings copied from the same work.

It is quite clear that these 'abnormalities' of the urodele limb express the same morphogenetic pattern as that responsible for the stegocephalian carpus and tarsus; and in both groups there is a marked tendency to lose the proximal post-centrale before the distal, as compared with the hypothetical ancestral condition.

In other living tetrapods, as clearly shown by the work of Steiner (1921, 1934, 1935), Holmgren (1933, 1939), and others, a general pattern involving a similar number of elements can be perceived, i.e. there are four centralia in addition to the precentrale. There is often much cause for argument in the determinations made by various workers, and a brief note is desirable. The anuran foot is highly specialized, but in the hand the radiale and precentrale ('element y') are best understood as preaxial branches of the axial row, turned over the distal end of the radius (figure 4G). The 'centralia 1 and 2' (Holmgren 1933) represent the main axis, 'centralia 3 and 4' being postcentralia. In the birds also there is great specialization, but in nearly all other tetrapods there is a very striking resemblance between the pattern of hand and foot of the same animal at early stages. In the Chelonia the digits are often markedly distinct from the splayed-out limb-stem. The descriptions of *Chrysemys* given by Holmgren (1933) and Steiner (1934) are in good agreement, but the element regarded as prepollex by Holmgren (figure 4H) is identified as radiale by Steiner. In *Sternotherus* (Holmgren 1933, figure 66) the so-called 'C 2' may really be a radiale or precentrale, because Holmgren's 'C 3' has the same relations as 'C 2' in other forms. In lizards without reduced limbs (Holmgren 1933; Steiner 1922, figure 9) the main axis (with intermedium and axial centralia) directed towards digits I-II is sometimes easily seen, though detailed interpretation is open to question in other cases (a very tentative reinterpretation of *Agama* is given in figure 4I). The crocodilian hand and foot also shows essentially the same pattern, with probably two post-centralia (as interpreted by Steiner 1934 and Holmgren 1933) and with the radiale or tibiale intimately connected with the main axis. However, the 'C 2' element of their nomenclature is medial (radial or ulnar) to the proximal end of digit I, and 'C 3' has the correct relationship for the

centrale distale; it is therefore possible that the former is really a precentrale, though in view of the structural modifications noted in some lizards mere topographical relations may be inadequate for proof. In *Mus* (Holmgren 1933) the

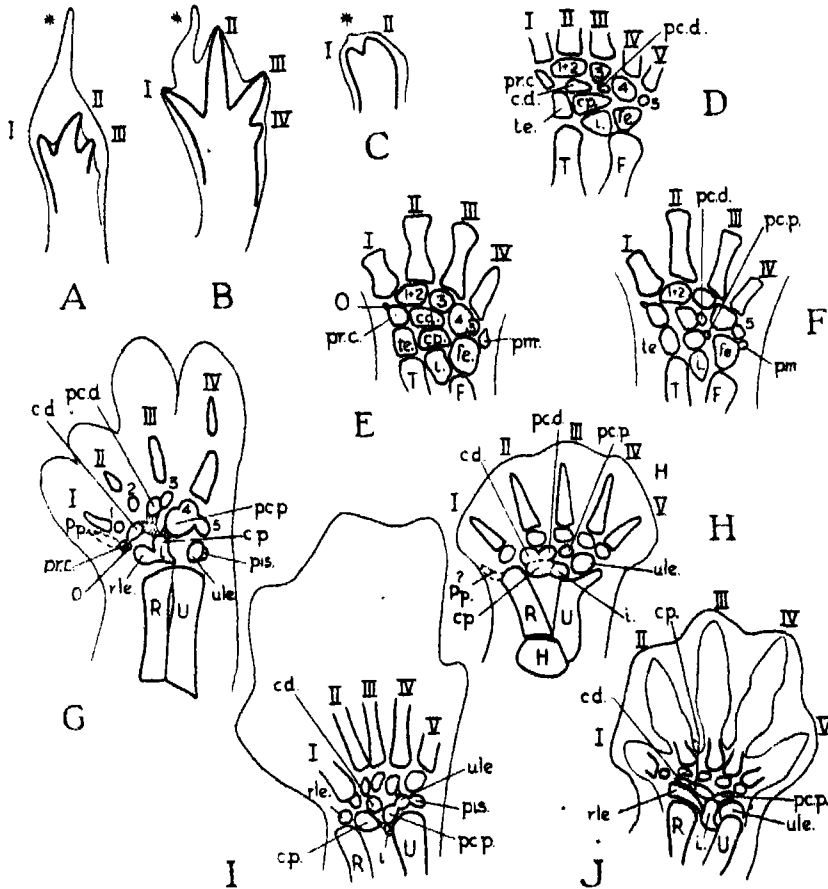


FIGURE 4. Limb structure in living tetrapods; developmental stages showing cartilaginous elements. Redrawn from Holmgren (1933). A, B, foot and hand of the urodele *Hynobius retardatus* showing position of 'fin-tip' between digits I and II. C, hand of the urodele *Cryptobranchus* showing similar structure. D, foot of the urodele *Megalobatrachus* (after Emery). Note postcentrale distale. E, F, normal and 'abnormal' feet of the urodele *Salamandrella* (after Schmalhausen). In the latter note separate postcentralia. G, hand of the anuran *Pelobates*. Note radiale and precentrale as branches of axial region. H, hand of the chelonian *Chrysemys marginata*. I, hand of the lacertilian *Agama colonorum*. J, hand of the mammal *Mus*. The nomenclature of the cartilage elements is modified from that of Holmgren in conformity with that used in this paper. (For explanation see legend to figure 1.)

digital blastemas are again markedly separate, and the main axis, curving round in front of the radius (tibia) towards digits I-II, is quite clearly seen, with the radiale as a preaxial branch (figure 4J). Holmgren's account offers one difficulty, his 'C 3' is intercalated between what should be (by his nomenclature) adjacent members of

the axial series. This may be true, because minor readjustments of pattern must be expected during adaptation to new structural conditions (cf. *Agama*, as interpreted by Holmgren 1933, figure 73), but it is rather probable that the element identified as fused 'intermedium and proximal centrale' may really be the intermedium alone, and that his 'C 3' is really the centrale proximale, so that the postcentrale distale would have disappeared or failed to differentiate from the central axis (figure 4J). Confirmation may be obtained from the foot, when the fibulare and postcentrale proximale have the same relationship to the intermedium as the corresponding elements have to the 'intermedium and proximal centrale' in the hand (cf. the close resemblance in the pattern of the skeletal anlagen of carpus and tarsus in general).

It may therefore be accepted that the general plan of the amniote limb which appears in the works of Steiner and Holmgren is (with possible minor exceptions) correctly established. In the present interpretation there may be recognised a series comprising an intermedium—centrale proximale—centrale distale, directed from the ulna towards digits I and II, 'preaxial' rays forming radiale and sometimes a precentrale; and postaxial elements forming the ulnare (fibulare) and sometimes two postcentralia. This pattern, though often greatly modified, is essentially the same as that of urodeles and Stegocephalia, contrary to the belief of Holmgren, and this pattern can be derived immediately from the rhipidistian paddle. The complete similarity in the structure of pectoral and pelvic limbs implies that the pectoral and pelvic fins of the ancestral rhipidistian fishes must have had similar numbers of mesomeres and rays.

The two theories outlined in figure 2A, A'' are both very simple, but the first, which is favoured by the writer, involves very little change of the rhipidistian paddle, and the addition only of structures which are clearly distinct in tetrapod ontogeny. It may be over-simplified, but the main points—the main axis extending towards digits I–II, and the separation of the ulnare (fibulare) and postcentralia from the main axis—are based on very substantial evidence. The nature of the radiale (tibiale) and precentrale is more dubious. Because the radiale (tibiale) and precentrale appear to branch from the axis in very many tetrapods—conditions not otherwise easily explicable—the first theory (figure 2A, or its modification figure 2A') is deliberately put forward in preference to the second (figure 2A''), in spite of the apparent contradiction provided by urodeles.

The first theory also explains satisfactorily many peculiarities of the stegocephalian limb; some of these are also corollaries of the second theory, and of that of Gregory & Raven (1941), which in part closely resembles the present schemes:

(i) The rotation of the proximal segment of the limb during the stride, shown for example in the screw-shaped glenoid articulation of the primitive tetrapod humerus (Watson 1917), is apparently directly comparable with the axial rotation of the rhipidistian paddle during its movement from the abducted to the retracted position (figure 1 and p. 377). A similar movement affects the femur. In the most primitive tetrapods the fore- and hind-limbs were essentially alike, with their

proximal elements directed laterally. In most tetrapods the elbow is directed backwards and the knee forwards, but this is a secondary modification.

(ii) The distal end of the primitive tetrapod humerus closely resembles that of the 'humerus' of rhipidistia in which the epicondylar processes, with the ulna articulating between them, and the radius articulating with the ectepicondylar side, can be recognized. The pelvic limb has not a well-developed entepicondyle, a difference which should afford an interesting problem in locomotion mechanics.

(iii) The ulna (fibula) is always shorter than the radius (tibia) in primitive tetrapods, as in rhipidistia.

(iv) As emphasized by Gregory & Raven, the proportions of the limb in more primitive tetrapods approach those in rhipidistia. The archepodium beyond the ulna (fibula) becomes greatly shortened in later forms.

(v) The combination of a fixed tread and the rotation of the proximal element of a rhipidistian paddle undergoing the movements of a tetrapod limb would produce precisely the same twisting of radius and ulna (or tibia and fibula) as that characteristic of the normal primitive tetrapod. This twisting is well shown by Schaeffer's (1941) figure 3, but in his figure the ulna should lie above and behind the radius in the right forelimb of the urodele, not below and in front of it.

(d) *The segmental nature of the rhipidistian paddle and the tetrapod limb*

The rhipidistian paired fin has often been compared with that of primitive sharks (e.g. *Cladoselache*, *Cladodus*, etc.; cf. Moy-Thomas 1936, 1939). Each consists of a series of cartilaginous rays or radials (ossified in rhipidistia) borne on a corresponding series of elements forming a metapterygial axis. The metapterygial axis in sharks seems to be composed merely of separate proximal (basal) elements, each basal and radial being related to a single myotome in the manner made familiar by Goodrich. The most primitive sharks show that the whole fin may have been a lateral keel; it became freed by a posterior embayment between the metapterygial axis and the body wall. It is, therefore, not unreasonable to suppose that a very primitive rhipidistian may also have had keel-like fins which became mobile in the same way. This receives some confirmation from consideration of the small number of segments involved in rhipidistian fins—five mesomeres (basals) and radials in the pectoral of *Eusthenopteron*, four in the pelvic, and four (? five) in the pectoral of *Sauripterus*, each mesomere and radial presumably related to one myotome.* In the shark fin there may be a few anterior and posterior myotomes involved to such a small extent that they do not contribute to the skeleton or skeletal muscle, and one would expect the same to be true in rhipidistia and in primitive tetrapods. Moreover, each myotome is motor innervated mainly by one spinal nerve, with

* Moy-Thomas (1936) has suggested that the paired fins and some of the median fins of crossopterygians are so similar that they were probably formed by the familiar method of basal concentration of radial elements. I am unable to see more than a superficial similarity between the patterns; in *Eusthenopteron*, where both paired and median fin skeletons are known, the patterns are quite different.

minor contributions from the two adjacent nerves. In all living tetrapods for which I have been able to get information the brachial plexus is drawn from never more than six spinal nerves, usually from five, sometimes from as few as three; and the cutaneous (sensory) innervation is related to about eight or fewer spinal nerves. The pectoral fin of *Eusthenopteron* would presumably have motor innervation from six or seven spinal nerves, and in view of the distal reduction of representatives of the rhipidistian paddle in the archepodium this evidence favours the recognition of the mesomere + radial as related to a single myotome bud in rhipidistia.

In living sharks there is a great range in the number of myotomes contributing to each concentrated paired fin (cf. Goodrich 1930), and normally the number in the pectorals differ from that in the pelvics. In Rhipidistia variation is present even in the few types known. *Eusthenopteron* has five mesomeres in the pectoral, four in the pelvic (the most 'distal' or posterior being rudimentary), and *Sauripterus* has apparently only four in the pectoral. As suggested above, the prototetrapod type probably had six (five fully developed and one rudimentary) in each fin. This is a further strong point favouring the unique and monophyletic origin of the known tetrapods (cf. Westoll 1943).

The tetrapod digits are apparently mostly 'postaxial', as emphasized by Gregory & Raven. There are two obvious possible sources of their material—by outgrowth from adjacent mesoderm of the archepodium, or by migration of material from elsewhere, presumably down the 'postaxial' margin of the developing limb. In the first case changes in the carpus or tarsus would be expected to be invariably associated with profound changes in the number and arrangement of digits, but this is by no means the case. The other possibility affords a direct explanation of some peculiarities in the segmental motor innervation of the muscles of the hand and foot, and is in good agreement with the observed conditions during ontogeny.

In the rhipidistian paddle there were probably flexor and extensor muscles moving each radial on its basal mesomere, and moving adjacent segmental structures on one another. Presumably the muscular derivatives of each myotome would be innervated from one spinal motor root, with probably minor contributions from the two adjacent spinal nerves. The muscles of each mesomere-radial unit should be supplied mainly from the spinal nerve cephalad to that supplying the next most distal unit. In tetrapods this condition must be greatly modified by the new structural immobility of the distal part of the archepodium and by its partial reduction, and by the probable migration of muscle attachments. I do not have access to any account of the *segmental* motor innervation of the limb muscles except in *Homo*, in which great modification is to be expected—particularly in the lower extremity, in consequence of the assumption of a bipedal gait. In the upper extremity the muscles may be grouped according to their attachments, as shown in Table 1*. The segmental motor innervation is taken from the 1938 edition of

* In order to avoid lengthy discussion of the homologies of the muscles of the human arm with those of less specialized mammals, and of the common matrices from which various muscles develop, the table is arranged simply on the basis of attachments of the muscles.

TABLE 1. SEGMENTAL MOTOR-INNervation OF MUSCLES OF ANTERIOR
EXTREMITY OF *Homo* (MOSTLY FROM GRAY)

	C						T
	3	4	5	6	7	8	1
GROUP I. Body to shoulder girdle:							
trapezius	3	4	.	.	.	.	.
rhomboides major, minor	.	.	5	.	.	.	.
levator scapulae	3	4	5	.	.	.	.
subclavius	.	.	5	(6)	.	.	.
serratus anterior	.	.	5	6	7	.	.
pectoralis minor	.	.	.	.	7	8	(1)
GROUP II. Body to upper arm:							
latissimus dorsi	.	.	.	6	7	8	.
pectoralis major	.	4	5	6	7	8	1
GROUP III. Shoulder-girdle to humerus or radius:							
deltoid	.	.	5	(6)	.	.	.
subscapularis	.	.	(5)	6	.	.	.
supraspinatus	.	.	5	(6)	.	.	.
infraspinatus	.	.	5	(6)	.	.	.
teres minor	.	.	5	.	.	.	.
teres major	.	.	(5)	6	.	.	.
coracobrachialis	.	.	.	.	7	.	.
GROUP IV. Humerus to radius:							
biceps brachialis	.	.	5	(6)	.	.	.
pronator teres	.	.	.	6	.	.	.
flexor carpi radialis (to base of metacarpals 2 and 3)	.	.	.	6	.	.	.
brachioradialis	.	.	5	6	.	.	.
supinator	.	.	5	6	.	.	.
GROUP V. Humerus or radius to ulna:							
brachialis	.	.	.	6	(7)	.	.
triceps	.	.	.	6	(7)	.	.
anconeus	.	.	.	.	7	8	.
pronator quadratus	.	.	.	6	7	.	.
GROUP VI. Humerus or antebrachium to Carpus or fingers, with exception of those in Group VII:							
A. Extensor carpi radialis							
longior and brevior	.	.	.	6	7	.	.
extensor digitorum	.	.	.	.	7	.	.
extensor digiti minimi	.	.	.	.	7	.	.
extensor carpi ulnaris	.	.	.	.	7	.	.
abductor pollicis longus	.	.	.	(6)	7	.	.
extensor pollicis brevis	.	.	.	.	7	.	.
extensor indicis	.	.	.	.	7	.	.
B. Palmaris longus							
flexor carpi ulnaris	.	.	.	(6?)	.	(8?)	.
flexor digitorum sublimis	.	.	.	.	7	8	(1)
flexor digitorum profundus	.	.	.	.	(7)	8	(1)
flexor pollicis longus	.	.	.	.	.	8	(1)
GROUP VII. Small muscles of hand							
	.	.	.	.	.	.	1

Gray's *Anatomy*; slight differences are shown in the latest edition of Cunningham (1942). In each case the small muscles of the hand are said to be innervated from other sources (cervicals 6, 7 and 8) besides the first thoracic, but the clinical evidence for the entire motor innervation through T 1 is apparently strong, and is here accepted. It would be very valuable to discover experimentally the segmental motor innervation of the less specialized limbs of lower tetrapods.

It would be expected, on the basis of the preceding argument, that the innervation of groups III and IV should be the same, since humerus and radius are presumably derived from one 'segmental' unit, and this is found to be the case. The ulna represents a more distal (i.e. posterior) segmental unit, and group V shows unmistakably a more caudad innervation. Group VI should represent muscles from the distal part of the archepodium which have acquired distal attachments to the digits, and again there is a more caudad innervation of the flexors, though the extensors have the same innervation as group V. Finally, group VII shows the most caudad innervation of all. Omitting the small muscles of the hand, there is only a range of about four segments indicated by the nerve supply to the muscles of groups III-VI, which may be compared with a probable range of six (seven or eight including minor contribution from adjacent spinal nerves) in the ancestral rhipidistian; this is in accord with expectations (above, pp. 383, 384). It also seems highly probable from this evidence that the digits may have formed from a 'post-axial' inflow of mesoderm in the developing limb-buds of the ancestral tetrapods.

The hindlimb is much more complicated in *Homo*, but a somewhat similar arrangement can be made out without difficulty. Full discussion would require, besides analysis of homologies by the usual methods, much further information concerning the segmental motor innervation of more primitive tetrapod types. Meanwhile it appears that the present theory is not contradicted by what is known of the varied segmental innervation of limb muscles, and in particular it explains the remarkable fact that all digits seem to have their motor innervation from the most caudad spinal nerves of the limb plexus.

Finally, since the dipnoan and coelacanth fins have been used by previous authors in the comparison with tetrapods, it seems desirable to comment on them. The coelacanths originated in the upper Middle or lower Upper Devonian; the writer has material of a fish, from the basal Upper Devonian of Canada, which has a coelacanth head and squamation but a body form and median fins showing more or less normal rhipidistian conditions. It seems reasonable to assume that the coelacanths are rhipidistian derivatives, and the structure of their pectoral and pelvic fins shows that the 'radials' are mostly lost, the fin skeleton consisting essentially of the mesomerall axis (cf. Stensiö 1932; Moy-Thomas 1937). The Dipnoi appear to have diverged, perhaps in the early Devonian, from the stock which gave rise also to the Rhipidistia. Early members show the paired fins comparatively more elongated and narrow than in osteolepids, and the narrowing becomes more obvious in later forms. In *Scaumenacia* there is some evidence that

the preaxial edge of the pectoral fin was directed medio-ventrally when the fin was in the trailing position along the ventral flank; this orientation is the same as that in rhipidistia. In the living *Epiceratodus* the preaxial margin of the pectoral is directed dorsally, perhaps a consequence of the much more dorsal position of the fin on the flank. The internal skeleton of the paired fins is not known in any early dipnoan. The '*Ceratodus*' type with its long series of axial elements, and both pre- and postaxial 'radials', bears to the rhipidistian paddle a somewhat similar relationship to that between the *Pleuracanthus* and cladodont shark-types, and the postaxial 'radials' may be regarded as secondary developments. The number of axial elements in the pectoral and pelvic fins together may, in well-grown living Dipnoi, exceed the number of pre-anal myomeres, and it is therefore highly probable that the long segmented axis is a secondary modification. This is an example of 'secondary polyisomerism' (Gregory's term). It is possible that such a fin was derived from a form with a comparatively short axis like that of Rhipidistia.

The somewhat similarly proportioned elongated fins of holoptychiids may be a parallel development, but their structure is not known.

My colleagues in the Department of Anatomy have been most kind in discussing with me some of the problems outlined above. This paper is largely dependent on the results of work done in the U.S.A. and Canada in 1937 with the aid of grants from the Royal Society and the J. B. Tyrrell Fund of the Geological Society of London, and of work on the Devonian fishes of Scotland aided by a grant from the Daniel Pidgeon Fund of the latter Society; for these grants, grateful acknowledgement is made.

REFERENCES*

- Böker, H. 1927 Ursprung der Extremitäten Wirbelt. *Z. Morph. Anthr.* 26, 1-58.
 Broom, R. 1913 Origin of the Cheiropterygium. *Bull. Amer. Mus. Nat. Hist.* 32, 459-464.
 Cunningham, 1942 *Text-book of anatomy*, 7th ed. Ed. Brash and Jamieson. Oxford University Press.
 Goodrich, E. S. 1901 Pelvic girdle of *Eusthenopteron*. *Quart. J. Micr. Sci.* 45, 311-324.
 Goodrich, E. S. 1930 *Studies on vertebrates*. London: Macmillan and Co.
 Gray, 1938 *Anatomy*, 27th ed. Ed. Johnston and Willis. London: Longmans, Green and Co.
 Gregory, W. K. 1915 Origin of the Tetrapoda. *Ann. N.Y. Acad. Sci.* 26, 317-383.
 Gregory, W. K. 1935 Pectoral girdle and fin of *Sauripterus*. *Proc. Amer. Phil. Soc.* 75, 873-890.
 Gregory, W. K., Miner, R. W. & Noble, G. K. 1923 The carpus of *Eryops*. *Bull. Amer. Mus. Nat. Hist.* 48, 279-288.
 Gregory, W. K. & Raven, H. C. 1941 Origin of paired fins and limbs. *Ann. N.Y. Acad. Sci.* 42, 273-360.
 Holmgren, N. 1933 Origin of tetrapod limb. *Acta Zool., Stockh.*, 14, 185-295.
 Holmgren, N. 1939 Origin of tetrapod limb. *Acta Zool., Stockh.*, 20, 89-124.
 Howell, A. B. 1935 The primitive carpus. *J. Morph.* 57, 105-112.
 Moy-Thomas, J. A. 1936 Evolution of pectoral fins and fore-limb. *Sch. Sci. Rev.* 17, 592-599.
 Moy-Thomas, J. A. 1937 Carboniferous coelacanth fishes. *Proc. Zool. Soc. Lond.*, B, 107 383-416.
 Moy-Thomas, J. A. 1939 Evolution of the Elasmobranchs. *Biol. Rev.* 14, 1-26.

* Titles of papers have been abbreviated.

- Petronievics, B. 1918 Pectoral fin of *Eusthenopteron*. *Ann. Mag. Nat. Hist.* 9) 2, 471-476.
- Romer, A. S. & Byrnes, F. 1931 The pes of *Diadectes*. *Palaeobiologica*, 4, 25-48.
- Schaeffer, B. 1941 Evolution of tarsus in amphibians and reptiles. *Bull. Amer. Mus. Nat. Hist.* 78, 395-472.
- Schmalhausen, J. J. 1910 Entwicklung der Extremit. von *Salamandrella*. *Anat. Anz.* 37, 431-446.
- Steiner, H. 1921 Hand und Fuss der Amphibien. *Anat. Anz.* 53, 513-542.
- Steiner, H. 1922 Entwicklung des Vogelflügelskelettes. *Acta Zool., Stockh.*, 3, 307-360.
- Steiner, H. 1934 Hand- und Fuss-Skel. bei den Crocodiliern. *Rev. Suisse Zool.* 41, 383-396.
- Steiner, H. 1935 Entwicklung des Chiropterygium. *Rev. Suisse Zool.* 42, 715-729.
- Stensiö, E. A. 1932 Triassic fishes. *Medd. Grönland*, 83, no. 3.
- Watson, D. M. S. 1913 On the primitive tetrapod limb. *Anat. Anz.* 44, 24-27.
- Watson, D. M. S. 1917 The evolution of the tetrapod shoulder girdle and fore-limb. *J. Anat. Lond.*, 52, 1-63.
- Westoll, T. S. 1943 Origin of the tetrapods. *Biol. Rev.* 18, 78-98.

The hyomandibular of *Eusthenopteron* and the tetrapod middle ear

BY T. S. WESTOLL, PH.D., D.Sc., F.G.S.

Department of Geology and Mineralogy, University of Aberdeen

(Communicated by D. M. S. Watson, F.R.S.—Received 18 February 1943.)

The hyomandibular of *Eusthenopteron* was incorrectly described by Sternberg: it compares closely with that of *Ectosteorhachis* described by Romer. Sternberg's account of the otico-occipital part of the cranium is also inaccurate, because his material is badly crushed. His phylogenetic conclusions are therefore suspect.

It is considered that Eaton's theory of the transformation of the Rhipidistian hyomandibular to the primitive tetrapod 'stapes' or columella is essentially correct. The processus opercularis becomes the processus tympanicus (extrastapes or extracolumella).

The tympanic and quadrate processes of the columella (stapes) are homologous in Stegocephalia and Reptilia.

The development of a freely movable neck-joint in early tetrapods was associated with considerable modifications in the hyobranchial apparatus, resulting in the loss of the operculum, allowing the formation of a primitive tympanic diverticulum, and leading also to the linear discontinuity of the hyomandibular and ceratohyal.

The living Amphibia possess two proximal 'middle ear' elements—a 'plectrum' and an 'operculum fenestrae ovalis'—not found separately in any amniote. They may be recognized as the homologues of processes of the stapes of fossil labyrinthodont Stegocephalia, from which group (including their small relatives the 'Phyllospondyli') the living Amphibia were probably derived. In the whole group the chorda tympani was primitively post-tympanic, and only a dorsal tympanic diverticulum was present.

In primitive Reptilia the region of the otic notch moved ventro-laterally. The dorsal diverticulum was still present but reduced in importance, and was supplemented by a recessus infrastapedialis which enveloped the quadrate process of the stapes.

In typical Theromorpha a further downward movement of the skeletal roof of the tympanic diverticulum took place, and an additional extension of the cavity (recessus mandibularis) was formed, applied to the lower inner surface of the mandible and later accommodated in a special cavity bounded by the reflected lamina of the angular.

In mammals these parts of the tympanic cavity have enveloped the posterior bones of the primitive mandible and the quadrate, and the old tympanic membrane has enlarged and migrated somewhat antero-ventrally. The relations of the mammalian tympanic membrane and cavity to skeletal elements, muscles and nerves are shown to be direct consequences of this theory.

Broom's suggestion that the cavity below the reflected lamina of the angular in *Theromorph* housed a salivary gland is discussed, and the possible position of the parotid and submandibular glands in cynodonts is indicated.

1. THE HYOMANDIBULAR OF *EUSTHENOPTERON*

The structure of the brain case of Rhipidistia has recently been described in *Rhizodopsis* (Säve-Söderbergh 1936) and *Ectosteorhachis* [*Megalichthys*] *nitidus* (Cope) (Romer 1937). Both investigators showed that the hyomandibular articulated with the brain case at two facets, one above and one below the jugular canal (transmitting the lateral head vein). This condition is of great theoretical importance in discussing the hyomandibular of fishes. More recently, Romer (1941) has given an excellent account of the hyomandibular of *Ectosteorhachis*, and has commented afresh on the brain case. Sternberg (1941) has also described the brain case and related structures of *Eusthenopteron*, in which, he asserts, the hyomandibular was different in its articulation with the brain case and relationship to the branches of truncus hyomandibularis VII. Material of *Eusthenopteron* in my possession shows that Sternberg's account is inaccurate and misleading in this and some other particulars. It seems desirable to publish a short account of the hyomandibular of this form and to comment briefly on the recent work of Romer and of Eaton (1939*a, b*) on the relationships of the hyomandibular to the operculum, and of the tetrapod stapes.

Available material of *Eusthenopteron* consists of several good skulls, showing some details of the hyomandibular, collected by Dr W. Graham-Smith and the writer from the type locality of Scaumenac (Escuminac) Bay, Bonaventure Co., Province of Quebec, and an isolated hyomandibular (on loan from the Geological Survey of Canada, no. 8691) from the same locality. The last mentioned had partly broken loose from the matrix during transit, and a thin layer of matrix was easily removed, allowing examination of both aspects of the antero-dorsal part of the bone. Although much flattened, this specimen is satisfactorily preserved and shows that the *Eusthenopteron* hyomandibular closely resembles that of *Ectosteorhachis*.

In general shape (figure 1) the bone corresponds perfectly well to Sternberg's figures. However, it is now clear that what Sternberg figures as the 'dorsal' view is really the lateral or antero-lateral aspect of the bone, and the 'ventral' view is really medial. The two articular heads of the bone are not well differentiated and the proximal end is flattened: between the heads, as in *Ectosteorhachis*, is a thin connecting lamina. The medial aspect of the proximal part of the bone shows, between the two heads, a very marked groove (figure 1, *t.hm.*) gradually passing distally into a tunnel (*f.p.*), which emerges on the lateral face (*f.d.*) just above the protuberance labelled 'processus opercularis' by Sternberg. This groove and tunnel are clearly for the main hyomandibular trunk of nerve VII. More distally the nerve

seems to have divided into the usual three rami. The almost parallel grooves (figure 1, *r.m.*, *r.h.*) for ramus mandibularis VII and ramus hyoideus VII are clearly shown on the distal part of the lateral face; the latter runs distally between sharp narrow ridges. Immediately behind and above the emergence of the tunnel, on the lateral face, is a steeply pitching hollow which I have been unable to prepare fully because of the fragility of the bone; it bears the characteristic minute cross-wrinkling of the impressions of the branches and trunks just mentioned and seems

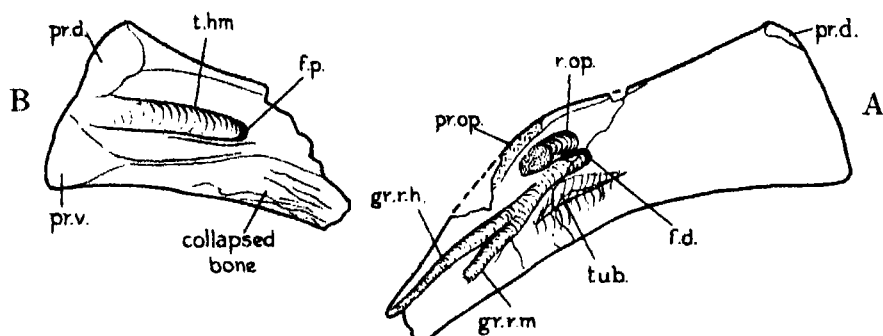


FIGURE 1. Hyomandibular of *Eusthenopteron foordi* Whit., Escuminac Formation, Escuminac Bay, P.Q., Canada, Geol. Surv. Canada, no. 8691. A, antero-lateral view. B, medial view of proximal end.

Explanation of lettering, figures 1-6: *a*, anterior process (cf. plectrum) of columella; *Ang.* angular; *Art.* articular; *C.* coronoid; *Ch.* ceratohyal; *Col.* columella auris; *cr.o.* crista obliqua (of pterygoid or columella); *c.ty.* chorda tympani; *D.* dentary; *e.a.m.* external auditory meatus; *E.O.* exoccipital; *e.t.* Eustachian tube; *f.d.*, *f.p.* distal and proximal openings of tunnel, for truncus hyomandibularis; *Gon.* goniale; *gr.r.h.* groove for ramus hyoideus; *gr.r.m.* groove for ramus mandibularis; *Hmd.* hyomandibular; *Inc.* incus; *I.Pa.* interparietal; *M.C.* Meckel's cartilage; *m.m.* manubrium mallei; *p.* posterior process (cf. operculum) of columella; *Pa.* parietal; *p.oc.* paroccipital process; *P.Pa.* postparietal; *PQ.* palatoquadrate; *Pr.Art.* prearticular; *pr.d.* processus dorsalis; *pr.h.* processus hyoideus; *pr.op.* processus opercularis; *pr.pt.* process for attachment of reptilian pterygoid muscle; *pr.qu.* processus quadratus; *pr.ty.* processus tympanicus; *pr.v.* processus ventralis; *ps.* reptilian pseudo-temporal muscle; *Pt.* pterygoid; *pt.f.* pterygoid flange; *pt.int.* mammalian pterygoideus internus muscle; *pt.r.* reptilian pterygoideus muscle; *Qu.* quadrate; *Qu.J.* quadratojugal; *rec.is.* recessus infraostapedialis; *rec.md.* recessus mandibularis; *r.h.* ramus hyoideus VII; *r.m.* ramus mandibularis VII; *r.op.* ramus opercularis VII; *r.p.* retroarticular process; *S.Ang.* surangular; *S.Hy.* 'stylohyal'; *S.O.* supraoccipital; *Sp.* splenial; *Sq.* squamosal; *St.* stapes; *S.Tem.* supratemporal; *Tab.* tabular; *t.ty.* tensor tympani muscle; *tub.* tuberosity on hyomandibular; *ty.* tympanum; *ty.d.* tympanic diverticulum; *Ty.R.* tympanic ring.

without doubt to be for ramus opercularis VII. This ramus has therefore a slightly different course from that in *Ectosteorhachis*, where it branches from the other parts of the nerve while still in the canal, but this difference is not important. Sternberg's figure 8b shows clearly the entry of the canal for the main hyomandibular trunk of nerve VII; his figure 8a shows the deep hollow above '*pr.op.*' into which the canal opens distally, and the other deep hollow ('dorsal sulcus') includes the steep foramen for ramus opercularis.

The dorsal edge of the bone is not very well preserved. The region of the slight angulation is thickened and seems to be very much roughened, and is continued proximally as a sharp narrow ridge near the dorso-lateral margin of the bone. The tuberosity called 'processus opercularis' by Sternberg is slightly roughened on its postero-dorso-lateral aspect, but this may be due to abrasion, since other specimens do not show this roughening.

The other structures shown on the medial surface of the bone as preserved do not correspond very closely with those of *Ectosteorhachis*, and it is difficult to decide how far they are due to the compression of a bone which seems to have consisted of a well-ossified 'shell' with loose or little ossification within. In the material available only the proximal part of the medial face is shown, and it is not possible to be sure of the presence of the groove, possibly for ramus pretrematicus IX, described by Romer in *Ectosteorhachis*, though it may well be present. The impression of the lateral head vein is much less definite than shown in Romer's figure.

The identification of the processus opercularis is not immediately clear. It seems very probable that the dorsal angulation played at least some part in the support of the operculum (cf. p. 398), and it has similar morphological relationships to those of the opercular articulation in *Ectosteorhachis*. The 'processus opercularis' of Sternberg's account may be properly so called, though it is more likely that it was related to other structures. The protractor hyomandibularis muscle, or (where this is not separately developed) a tendon from the corresponding part of the levator arcus palatini musculature, is often attached to the hyomandibular in a similar position in fishes. The *Ectosteorhachis* condition could be attained by a dorsal or dorso-lateral extension of the 'processus opercularis' of Sternberg's account so as to cover still more of the canal for the seventh nerve and to meet the thickening at the dorsal angulation. The canal for ramus opercularis VII would then be more or less invisible in lateral view.

The orientation of the whole hyomandibular has been somewhat misinterpreted by Sternberg. In the living animal the cheek plates were rather steeply inclined downward, from the skull roof, and the hyomandibular ran at first almost posteriorly, and then postero-ventrally and slightly laterally, from the cranium. The flattening of the skull after death usually involves the disintegration of the gular region and the lateral splay of the palato-quadrates and cheek bones. The hyomandibular is also involved and is thus detached from the cranium. The ventral head of the bone, being longer than the dorsal head, may still lie apparently close to the cranium and give a false idea of natural articulation. Sternberg shows (his figure 2) possible areas of articulation for the hyomandibular on the endocranium of *Eusthenopteron*, both dorsal to the jugular canal. The facet for the 'posterior' (dorsal) head is also labelled 'articular surface for the first branchial arch' in his figure 1. I find Sternberg's account of the otico-occipital part of the brain case extremely difficult to follow on my material, which I have not now leisure to prepare properly. Meanwhile I have no doubt that Sternberg's material is strongly crushed, and suspect that collapse effects are responsible for some of the peculiar conditions

he describes. Examination of his stereo-photographs goes far to confirm the suspicion. The occipital region as restored by him is altogether too shallow (cf. Bryant 1919, plate 14, figures 2-5). It follows that Sternberg's phylogenetic and comparative anatomical conclusions are suspect.* Dr Jarvik of Stockholm has been engaged in a monograph on *Eusthenopteron* which should provide a solution to these difficulties.

The *Eusthenopteron* hyomandibular is therefore directly comparable with that of *Ectosteorhachis* and was almost certainly articulated with the endocranium in the same fashion. Sternberg's work takes no account of the changes during post-mortem flattening of the skull, and by insufficient preparation of the hyomandibular he has overlooked important characters.

The five main attachments of the rhipidistian hyomandibular may be named as follows: two proximal attachments, the processus ventralis to the wall of the otic capsule in the saccular-lagenar region, and the processus dorsalis reaching up almost to the dermal skull roof; the processus opercularis extending to the bony operculum; the processus quadratus to the quadrate; and the processus hyoideus to the ceratohyal.

2. THE TETRAPOD COLUMELLA AURIS AND STAPES

Considerable contributions to our understanding of the nature and evolution of the auditory apparatus in primitive tetrapods have recently been made by Bystrow (1938, 1939), Bystrow & Efremov (1940), Eaton (1939*a, b*), Romer (1941), and others, but there still remain many interesting problems which have received little attention from palaeontologists since Sushkin's paper of 1927. For some of these the material for a partial solution seems to be available; these concern the nature and evolution of the tympanic chamber and its membrane, and the relations of these structures to the visceral-arch skeleton and the branches of the truncus hyomandibularis VII.

Eaton (1939*a, b*) has made the suggestion that the part of the columella auris or stapes inserted in the tympanum—i.e. the extracolumella or extrastapes, which I propose to call the processus tympanicus—is homologous with the processus opercularis of the crossopterygian hyomandibular, a homology that had been independently worked out by the writer apparently on other grounds. Eaton considers that the crossopterygian hyomandibular had five attachments, which have been named and discussed above. It is possible that another attachment should be considered—namely, the tuberosity, identified by Sternberg as 'processus opercularis', on the hyomandibular of *Eusthenopteron*; this may be for a tendon from the protractor hyomandibularis or the levator arcus palatini. On the whole it is improbable that this process is as important as the others, it is not represented in *Ectosteorhachis*.

* I have material of a proto-coelacanth from the Escuminac formation, still retaining many rhipidistian characters of the body but with the jaws showing characteristic coelacanth modifications. It differs strikingly from *Eusthenopteron*.

The nature of the hyoid and mandibular skeleton in Rhipidistia is fairly well known in broad outlines. The palato-quadrate, covered internally by a large pterygoid bone, runs back almost horizontally, close to the ventral lamina of the temporal region, from just above the basipterygoid process to about the level of the hyomandibular attachment, behind which the upper margin curves down to the quadrate region. The hyomandibular lies very close to the upper part of the palatopterygoid complex, as shown in figure 2. The details of the distal unossified parts of the hyomandibular are unknown, but it is certain that there was a connexion with the ceratohyal, and very probably another with the quadrate. In available rhipidistian material I have not found any 'epihyal' ('stylohyal' or 'interhyal') or 'symplectic' elements; there seems to be no room for the latter to be developed.* The ceratohyal runs rather close inside the back of the mandible.

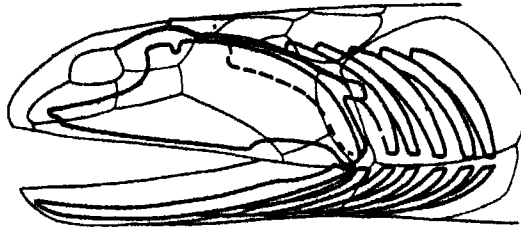


FIGURE 2. Probable arrangement of visceral-arch skeleton in an Osteolepid. The branchial arches shown schematically. Dermal skull as though transparent, after conditions in *Osteolepis* (cf. Westoll 1943).

The first branchial arch follows close on the hyoid in Rhipidistia. The apposition of the hyomandibular and quadrate region confirms the identification of the groove on the hyomandibular (Romer 1941) as for ramus mandibularis VII, and makes it fairly clear that no pharyngeal diverticulum of any size comparable to a tympanic chamber could have been developed. The processus opercularis articulated with the opercular near its anterodorsal corner; in *Eusthenopteron* there is a shallow pit on the inner face of the opercular which may have been occupied by cartilage for the articulation, exactly as in *Polypterus* (Allis 1923, plate 11, figure 33). This pit cannot possibly have articulated with the tuberosity on the hyomandibular of *Eusthenopteron*, which was labelled 'processus opercularis' by Sternberg, but could have been in articulation with a small cartilage extension at the angle of the bone in the position of the processus opercularis of *Ectosteorhachis*. The processus opercularis therefore lies behind the squamosal, about the level of the top of the preopercular, below and behind the hinder margin of the homologue of the tetrapod tabular, and lateral to the paroccipital process. By removal of the opercular bones and

* Symplectic and 'stylohyal' cartilages are known in the living coelacanth *Latimeria* (Smith 1939). The former is characteristic of bony fishes in which the suspensorium is vertical or forwardly inclined; great development of both elements may be associated with this forward shift. They seem to represent separate centres of chondrification in the quadrate and ceratohyal processes or ligaments from the hyomandibular.

forward movement of the quadrates, the processus opercularis would be left in a notch bounded by the tabular and paroccipital process above, and the squamosal and preopercular (with the quadrate) in front. It has therefore all the morphological relationships of the part of the stegocephalian columella or stapes (extracolumella or extrastapes) directed to the otic notch. In Stegocephalia too there are indications of the other four attachments—the processus ventralis to the fenestra (or pseudofenestra) ovalis, the processus dorsalis (often reduced) to the crista parotica or the tabular, the processus quadratus (marked only, in most Stegocephalia, by a pit or tubercle on the quadrate for attachment of a cartilaginous or ligamentous connexion), and the processus hyoideus which, at some stage of development, is a morphological necessity and which may have been degenerate or ligamentous in the adult.

At first sight the stapes of a primitive reptile (e.g. *Captorhinus*—Sushkin 1927, 1928; Price 1935; and cf. figure 5B) is so different as to suggest quite other homologies. The main axis of the stapes runs postero-laterally, and the connexion to the quadrate is obviously the most important distal one. However, the distal end of the bone is not usually preserved, and a short extrastapedial projection may well have projected to the region of the tympanum, as in the cotylosaur *Milleretta* and the therapsids described by Broom (1936, 1937, 1938). These conditions are related to the extraordinary downward spreading of the paroccipital plate in these reptiles and their derivatives, correlated with a downward drag of the tabular (and supratemporal) and complete obliteration of the original opercular or 'otic' notch. The stapes and its processes seem therefore to be strictly homologous in the two groups, and in particular the extrastapes has the same morphological relations. The conditions in *Lacerta* confirm this so far as they are relevant, and all five processes are present. In all these tetrapod forms, truncus hyomandibularis VII presumably emerged from the cranium in front of the articulations of the stapes, passed backwards between the two proximal heads, then outwards and downwards, dividing into the hyoid ramus (which continues along the outer side of the ceratohyal) and the ramus mandibularis internus (chorda tympani). The ramus mandibularis externus (ramus mentalis) would probably only be preserved as an important nerve in those tetrapods with functional latero-sensory organs. In *Lacerta* the chorda tympani passes forward, over the extrastapes (processus tympanicus) and lateral to the processus quadratus of the stapes, to the inside of the mandible. The chorda tympani runs essentially in front of or through the tympanic diverticulum in all amniotes.

The living Amphibia present rather markedly different conditions, due in part to the degenerate nature of living Urodela and Gymnophiona, and the complex metamorphosis of Anura, but in part indicative of special properties. Alone among known tetrapods these groups have the fenestra ovalis partly plugged by two cartilages each presumably part of processus ventralis*—an anterior 'plectrum'

* According to several investigators the operculum fenestrae ovalis may be derived from the wall of the otic capsule. The nature of the 'stapes' of Stegocephalia supports Goodrich's view (1930) that it is of hyoid arch origin.

usually meeting the anterior border of the fenestra ovalis, and drawn out into a stylus ('columella') in some forms, and a posterior 'operculum' (here called 'operculum fenestrae ovalis' to distinguish it from the piscine operculum); a process may run to the paroccipital process (processus dorsalis); and there may also be contact with the quadrate. In Urodela the tympanum is not developed and the whole columella auris may be vestigial. In Gymnophiona the plectrum may be pierced by the stapedia artery, and runs to the quadrate; no tympanum is developed. In the only group in which a tympanum occurs—Anura—the chorda tympani is post-tympanic, in sharp distinction to amniotes.

The immediate problems therefore are two:

(1) How did the rhipidistian condition evolve into that characteristic of generalized tetrapods?

(2) Are the living Amphibia descended from early tetrapod ancestors with an otic structure different from that of Amniota in some details?

The discussion of some of these problems leads to the emergence of a third:

(3) How did the angular of mammal-like reptiles come to be so closely associated with the tympanic membrane that it could take over its support in the mammals?

3. THE EVOLUTION OF THE TYMPANIC CAVITY IN TETRAPODS

The key to this problem is that the Rhipidistia show the hyoid arch in close spatial relationship with both the mandibular and first branchial arches. The earliest tetrapods allow only a partial knowledge of the region, but it seems fairly certain that the line of the 'suspensorium' defined by the basal articulation and the quadrate remains extremely oblique, directed postero-laterally and somewhat ventrally; in comparison with the osteolepids, however, the quadrates are moved anteriorly. In spite of this apparent similarity of attitude, and the similar parallelism of the now modified hyomandibular to the postero-dorsal margin of the palatopterygoid complex, there has been a considerable change in the relationships of the rest of the hyoid arch and of the branchial arches to the skull and palate. The branchial arches of phyllospondyls and of *Dvinosaurus* (Bystrow 1938, 1939) are some little way behind the skull, and apparently no elements more dorsal than ceratobranchials are ever present in known tetrapods. These facts are to be correlated with the development of a neck-joint and the musculature for moving the head. The result is quite clear; while the hyomandibular retains its connexion with the cranium and quadrate, the ceratohyal is found relatively farther back, so that a marked gap has appeared between the hyoid and mandibular arches just behind the angle of the jaws. Moreover, the forward movement of the quadrates has allowed greater separation dorsally between the hyomandibular and the quadrate. This is an important change from the fish condition, and seems to have afforded an opportunity for the outward growth and expansion of the pharyngeal pouch between the mandibular and hyoid arches, in the region behind and above the processus quadratus (figure 3B). In addition there must have arisen the possi-

bility of expansion of this pharyngeal pouch behind and immediately above the angle of the jaw, between the processus quadratus and processus hyoideus (figure 3E). One or both of these diverticula tended to envelop the homologue of the hyomandibular, as they do in the ontogeny of living tetrapods. In this way the hyomandibular (becoming the columella auris or stapes) would come to lie partly within the tympanic cavity, so that the processus opercularis could take up its new function as the processus tympanicus, imbedded in the thin membrane where the

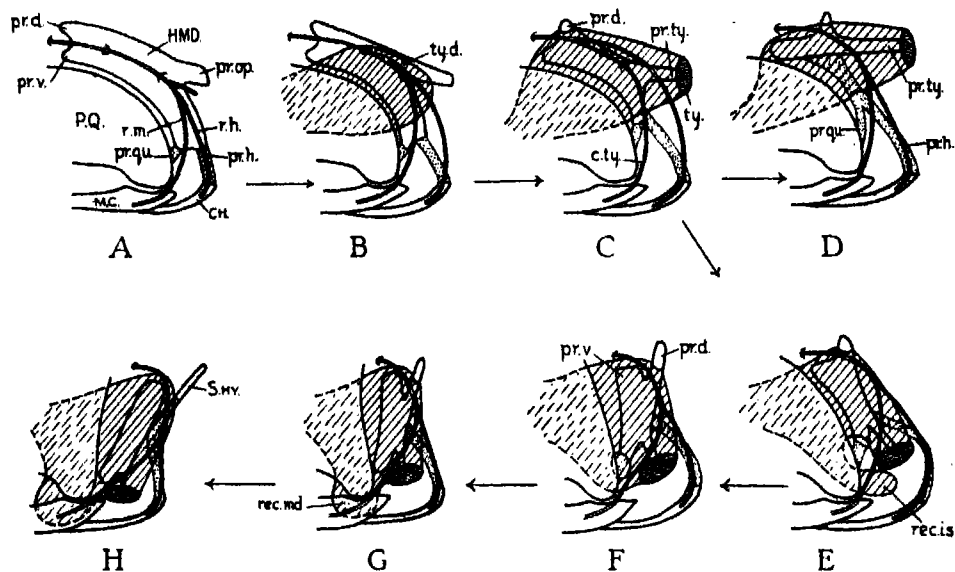


FIGURE 3. Diagrams showing mutual relations of hyoid and mandibular arch elements, tympanic diverticula and truncus hyomandibularis VII. A, osteolepid condition. B, development of recessus dorsalis tympanicus. C, possible conditions in primitive Amphibia. D, conditions in Stegocephalia and Anura. E, early reptilian condition; downturned of stapes, restriction of recessus dorsalis, development of recessus infrastapedialis tympanicus. F, conditions in many living reptiles; recessus infrastapedialis joins recessus dorsalis. G, conditions in primitive mammal-like reptiles; appearance of recessus mandibularis; dorsal process of stapes still prominent. H, conditions in advanced mammal-like reptiles; separation of dorsal process of stapes and processus hyoideus suggested. (For explanation see legend of figure 1.)

tympanic cavity approached the external morphological groove (the hyomandibular cleft) between the hyoid and mandibular arches. This change is easier to understand if Eaton's arguments (1939a) that the operculum of bony fishes consists partly of mandibular-arch tissue be accepted, for then the processus opercularis of Rhipidistia would have lain almost adjacent to the morphological hyomandibular cleft, and would have reached across to meet mandibular-arch material.

With the reduction in the dorsal part of the branchial region, and the increasing distance of the first branchial arch from the head, the rhipidistian derivatives would probably have found a bony operculum disadvantageous, particularly

during lateral turning of the head. Moreover, in larval living Amphibia, in phyllospondyls (Bulman & Whittard 1926) and in *Dvinosaurus* (Bystrow 1938, 1939) the most anterior branchial cleft functioning as a gill slit is between the first and second branchial arches, and only ceratobranchials are present. The loss of the opercular ossifications and the freeing of the processus opercularis may be understood in this light.

This series of changes affords a good working hypothesis, giving an adequate background for understanding the development of the tympanic diverticulum, the loss of the opercular bones, the break in some forms of the normal simple continuity of the hyoid arch usual in fishes, and the transformation of the processus opercularis into the processus tympanicus (cf. figure 3A-C). The elucidation of other points in the comparative anatomy of the middle ear require more elaborate treatment. Meanwhile it is desirable to refer to the view held by some workers that the tympanic diverticulum is derived from two pharyngeal pouches instead of from the single hyomandibular pouch. In *Homo*, for example, it is usually stated that the diverticulum is bounded posteriorly, in part, by the first branchial arch. But on comparative grounds there is much evidence that only one pouch is involved in other tetrapods. The difficulty seems to be one of nomenclature rather than of fact; if the diverticulum or diverticula from the hyomandibular pouch were to envelop the remains of the hyomandibular, as suggested above, the results might be quite similar to those consequent on the two-pouch theory.

The statement that the tympanum of tetrapods is derived from the spiracle of fishes is still encountered in some modern texts. Goodrich and others have clearly disposed of this view by studying the fate of the epibranchial placode which marks the true dorsal end of the hyomandibular slit, and which is considerably removed from the rudiments of the developing tympanic diverticulum.

4. THE AMPHIBIAN MIDDLE EAR

It was noted above that normal Stegocephalia present a markedly different aspect of stapes and 'otic notch' from that seen in typical reptiles; Anthracosauria and Seymouriamorpha show the former pattern. It is widely accepted that Anura are closely related to Stegocephalia, being perhaps derived from small types often classified as 'Phyllospondyli' (Watson 1940; Romer 1939). It is also increasingly clear (Pusey 1938, p. 538; Bystrow 1939; Westoll 1943) that there is little foundation for the suggestion of Säve-Söderbergh, Holmgren and others that the Urodela have been derived quite independently of the other Tetrapoda, perhaps from Dipnoi. A tympanic diverticulum is present only in Anura among living Amphibia; it is very dorsally placed, and the ramus hyomandibularis facialis passes back between the two cranial articulations of the columella, and lies post-tympanic. In Gymnophiona, where the columella is usually stout and directed towards the quadrate, the ramus hyomandibularis is morphologically post-columellar: no processus dorsalis is developed. In living Urodela the conditions are more com-

plicated. No tympanic diverticulum is present, and the stylus of the columella auris has a ligamental connexion with the squamosal or quadrate. The ramus hyomandibularis passes either ventrally to the columella in its entirety, or (in *Necturus*, etc.) the alveolar (including internal mandibular) branch runs ventrally and the jugular (hyoid) branch dorsally to the columella. These conditions can, at least in part, be explained by reference to fossil Stegocephalia, in which it is possible to observe conditions of the proximal part of the columella or stapes much resembling those seen in living Amphibia.

In typical Stegocephalia, such as *Eryops* (Sushkin 1927), *Edops* (Romer & Witter 1942), *Benthosuchus* (Bystrow & Efremov 1940), *Trimerorhachis* (Case 1935), *Buettneria* (Case 1931; Wilson 1941) and the Triassic forms described by Säve-Söderbergh (1936), the 'stapes' is elongated and is directed from the fenestra ovalis towards the otic notch, the relief of which strongly suggests that it supported a tympanic membrane (cf. particularly Bystrow & Efremov 1940). Near the distal termination, in at least some forms (*Eryops*, Sushkin 1927; *Edops*, Romer 1941, p. 151 and Romer & Witter 1942) there is a dorsal process meeting the tabular or crista parotica, and almost certainly homologous with the processus dorsalis. There is usually a roughened tubercle (tuberculum supratrochleare) on the quadrate, which probably marks the attachment of a ligament from the shank of the stapes (processus quadratus), and a branch of this ligament probably passed to the ceratohyal (processus hyoideus). Thus all five of the primitive processes of the rhipidistian hyomandibular can be recognized. Proximally, in several forms (e.g. *Benthosuchus*, *Buettneria*, *Trimerorhachis*, *Lyrocephalus*, *Edops*, and *Aphaneramma*), the stapes shows interesting features that have been insufficiently appreciated (figure 4). The anterior margin of the proximal face (processus ventralis) appears in many of these forms to have met the otic capsule just in front of the fenestra ovalis, and its antero-ventral margin may approach the parasphenoid and pterygoid, or may even be sutured against the parafenestral crista formed by those bones. From this anterior process the main axis of the stapes seems to run out to the otic notch. Behind it the proximal part of the stapes curves or flares out to a process which was obviously inserted in the fenestra ovalis; these parts are very clearly distinct in *Aphaneramma* (Säve-Söderbergh 1936, text-figure 47 and figure 4A of this paper) and in *Buettneria* (Wilson 1941 and figure 4B of this paper).^{*} These conditions cannot be matched in amniotes, and it seems very reasonable to suppose that the posterior process is homologous with the operculum fenestrae ovalis of modern Amphibia, fused with, or not yet separated from, the more an-

^{*} There is some confusion in Bystrow & Efremov's description of the columella of *Benthosuchus* (cf. figure 4E, D of this paper). In figure 15 the crista obliqua columellae is shown on the anterior edge, and a similar orientation is shown in figure 12. But in the text (pp. 26, 126) this crista is said to be visible in posterior view, and the 'processus basalis' of their description would thus be anterior. The latter orientation affords close comparison with other labyrinthodont columellae (particularly with *Edops*); perhaps the isolated *Benthosuchus* columella is the left, not the right. In that case the crista perhaps marks the attachment of the ligaments to the quadrate and ceratohyal.

terior 'plectrum' region. This seems to be strong evidence that Urodela, Gymnophiona and Anura were all derived from ancestral forms showing this type of stapes, and so far as is known those forms include only the labyrinthodont stegocephalians and probably their small ('phyllospondyl') relatives.

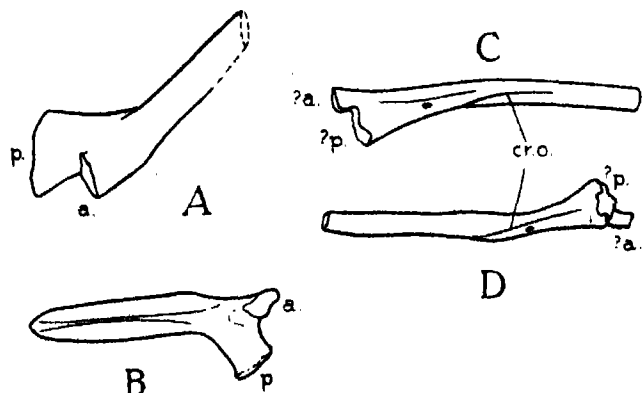


FIGURE 4. The ossified columella auris (stapes) in three Stegocephalia showing double nature of processus ventralis. A, *Aphaneramma* sp., anterior view of left columella, after Sæve-Söderbergh (1936). B, *Buettneria perfecta* Case, dorsal view of left stapes, after Wilson (1941). C, D, *Benthosuchus sushkini* (Efr.), after Bystrow & Efremov 1940. C, said to be ventral, D, anterior view of right stapes, but may be respectively ventral and posterior views of left stapes, and here labelled accordingly. a. anterior process, apparently corresponding to the plectrum of living amphibians; p. posterior process (? = operculum fenestrae ovalis).

The conditions in some of these forms allow us also to reconstruct much of the tympanic chamber, and with it the probable courses of the truncus hyomandibularis and some of its branches. This can best be done in *Benthosuchus* (Bystrow & Efremov 1940), in which the relief of the pterygoid and the otic notch allow almost complete certainty (figure 5A). From the columella (or stapes) to the quadrate and ceratohyal there passed connexions which were probably ligamentous. The ligament to the quadrate was almost certainly attached to the tuberculum supratrochleare; that to the ceratohyal probably diverged from the ligamentum quadrato-columellare not far above this point. The Eustachian tube passed through the level of the bony palate just lateral to the posterior end of the suture between the pterygoid and the parasphenoid, and it seems that Bystrow & Efremov were thoroughly justified in identifying a concavity on the postero-median face of the lamina ascendens of the pterygoid as 'excavatio tympanica' (but see Wilson 1941 for another view). This excavation is sharply bounded ventrally by the 'crista obliqua' on the pterygoid, to which Bystrow & Efremov suggest the margin of the musculus depressor mandibulae was attached; it seems also to have marked the effective antero-ventro-lateral extent of the tympanic diverticulum. There is also a 'recessus tympanicus' in the anterior face of the exoccipital (just behind the proximal end of the columella) which seems to have housed the posterior part of the proximal end of the diverticulum. The antero-median support of the tympanic

membrane seems quite certainly to have been on the tabular and squamosal; the relief of the ventral surface of these bones in the otic notch is striking. On the lateral part of the posterior face of the quadrate ramus of the pterygoid the crista obliqua bends suddenly down and continues in a direction slightly median of the tuberculum supratrochleare of the quadrate; I regard this as the course of the ligamentous connexion between the columella and the quadrate (and ceratohyal). An almost identical arrangement of ligaments in *Dvinosaurus* is restored by Bystrow (1938, figure 5); in this form the proximal (i.e. dorsal) part of the connexion

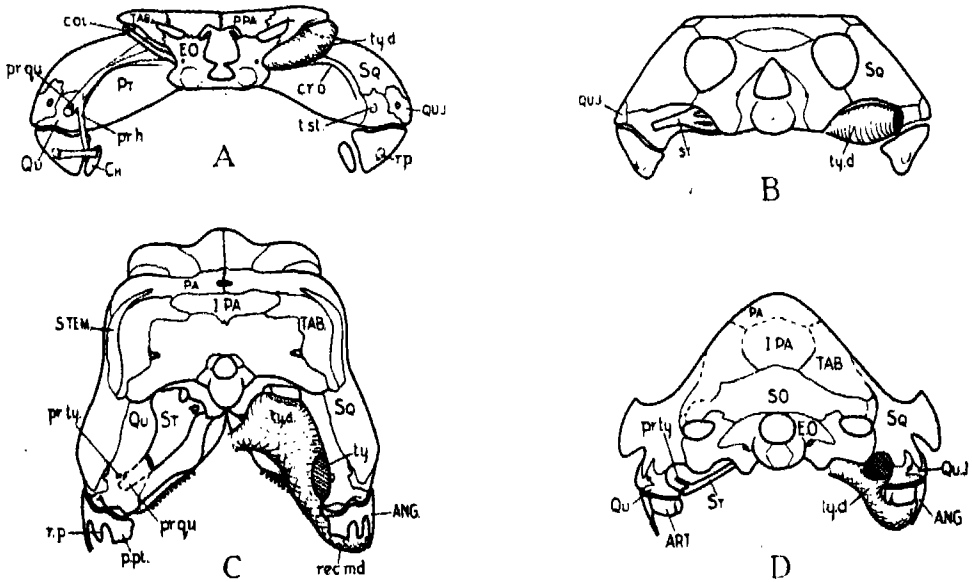


FIGURE 5. Attempted reconstructions (schematic) of tympanic diverticula in fossil tetrapods. A, the stegoccephalian *Benthosuchus*, adapted from Bystrow & Efremov. B, the cotylosaur *Captorhinus*, adapted from Sushkin, Price and others. C, the pelycosaur *Dimetrodon*, adapted from Romer & Price. D, the therocephalian *Lyledops*, adapted from Broom. In each figure the columella auris or stapes is shown on the left side, the tympanic diverticulum on the right. (For explanation see legend of figure 1.)

is actually an ossified process of the stapes. This reconstruction, and the similar one for *Benthosuchus* (Bystrow 1939, figure 15) would allow the depressor mandibulae musculature to reach the retroarticular process quite directly, and seem to be very highly probable. But if this is the case, no part of the functional tympanic diverticulum can have existed ventral to the ligaments to the quadrate and ceratohyal; the whole cavity must have been a very dorsal one, as in frogs. Moreover, the probable course of the truncus hyomandibularis of nerve VII can be indicated with some certainty. After its emergence from the cranium it must have branched off from the other trunci facialis, run posteriorly dorsal to the fenestra ovalis and thus over the proximal part of the columella but medio-ventral to the processus dorsalis, and divided into its mandibular and hyoid branches probably some

distance postero-laterally. The hyoid ramus would have continued along the ceratohyal. The mandibular ramus subdivided into the ramus mandibularis externus (ramus mentalis) (passing through the 'foramen paraquadratum' or quadrate foramen, here enclosed by the quadratojugal), which served the latero-sensory system in the mandible, and the ramus mandibularis internus or chorda tympani, which passed down to enter the inner face of the mandible between the (articular + surangular) and the prearticular. The chorda tympani passed to the mandible morphologically posterior and ventral to the tuberculum supratrochleare and processus quadratus (compare *Lacerta*, etc.). This account is not in full agreement with that of Bystrow & Efremov, who figure the entire truncus hyomandibularis passing below the columella, presumably by comparison with the highly specialized and probably degenerate conditions in living Urodela; but it is thoroughly self-consistent and in agreement with the conditions in other forms, especially in *Edops*, where Romer and Witter have shown the course of the truncus.

The conditions in most living Amphibia can be understood as modifications of this pattern. In Anura the columella retains the two proximal articulations—the processus ventralis to the region of the fenestra ovalis, the processus dorsalis to the crista parotica (cf. conditions in *Eryops*, *Edops*, *Dvinosaurus*, etc.). The truncus hyomandibularis facialis passes back between these two heads, as in nearly all other tetrapods and in Rhipidistia. Little is known of the variations in the detailed relationships of the columella to the quadrate and ceratohyal and the relations of the branches of the truncus hyomandibularis to these structures. The separate operculum fenestrae ovalis and plectrum recall the two ventral articulations of the stapes in some Stegocephalia. The extracolumella is embedded in the tympanic membrane and is clearly homologous with the processus tympanicus and the tip of the columella of such Stegocephalia as *Benthosuchus*. The Eustachian tube is dorso-laterally directed, and the tympanic chamber is very dorsal in position. I suggest as a working hypothesis that the annulus tympanicus of Anura may be the secondarily cartilaginous representative of the tabular of Stegocephalia, freed from the skull roof by the modification of the occipital surface: it retains the typical connexions with the crista parotica and the squamosal.

The columella of *Gymnophiona* has, in the case of *Ichthyophis* (Goodrich 1930, figure 508), a shape recalling that of some Stegocephalia, for at the stage figured the plectrum and operculum fenestrae ovalis are not separate, and the latter form a posterior process from the whole columella. The columella meets the quadrate; this is probably a secondary condition due to the failure of the tympanic diverticulum, and the distal end of the columella may well include the homologues of the processus tympanicus as well as of the processus quadratus.

The auditory region in Urodela is very peculiar (good brief review in Goodrich 1930). The auditory ossicles consist of separate plectrum ('columella') and operculum; the former is produced to a short process (stylus) and may be connected by ligament to the squamosal or quadrate. The ramus hyomandibularis passes usually ventral to the stylus and the ligamentum squamoso-columellaris, but

sometimes the jugular ramus (hyoid) passes dorsally over the stylus. It is exceedingly difficult to account for these conditions on any probable assumptions concerning the homology of the stylus and ligamentum squamoso-columellaris. But if one branch of the nerve can have its course above or below the stylus there is some justification for thinking that both branches may once have passed dorsal to the stylus. In that case the tip of the stylus may represent the greatly reduced processus tympanicus. During the evolution of Urodela the tympanic chamber has presumably failed, and it seems possible that the stylus has come to originate so late in ontogeny that the nerve has had the opportunity to readjust its course before the differentiation of the columella. In several Urodela the stylus may fail completely.

To sum up, the living Amphibia possess a common feature in the separate 'opercular' and 'plectral' connexions between the region of the fenestra ovalis and the proximal end (processus ventralis) of the hyoid arch; this can be compared only with labyrinthodont Stegocephalia (and probably with their small relatives or derivatives, the Phyllospondyli) among other tetrapods. The tympanic diverticulum in Anura and in the labyrinthodonts is or was dorso-laterally directed and is, or seems to have been, entirely dorsal to the distal part of the processus quadratus of the columella (or stapes). The conditions in Gymnophiona and Urodela may be understood as due to degeneration of the tympanic diverticulum. It therefore seems highly probable that all three groups of living Amphibia were derived from the extinct labyrinthodont-phylospondyl stock, contrary to the views of Holmgren and Säve-Söderbergh. Bystrow (1938) and Bystrow & Efremov (1940) have shown that the detailed anatomy of several labyrinthodonts can be compared closely with that of urodeles; a similarly detailed anuran comparison has been made by Säve-Söderbergh (1936).

5. THE MIDDLE EAR OF PRIMITIVE REPTILES

The dorsally directed stegocephalian columella is at first sight different from the more laterally directed reptilian stapes, but as shown above, this difference is correlated with the downturning of the region corresponding to the stegocephalian otic notch in reptiles. All the very early tetrapod types show conditions more or less comparable, in broad outlines, with the Stegocephalia, but forms like *Otocratia* (Watson 1926, and Westoll 1942) show how the lateral and ventral migration of the region corresponding to the stegocephalian otic notch could come about. I have elsewhere suggested that *Otocratia* may be related to the ancestors of some early reptiles, and on the other hand it is highly probable that *Otocratia* is a direct derivative of ichthyostegid-like forms, in which the otic notch had a general similarity to that of labyrinthodonts.

Full discussion of the reptilian middle ear is quite outside the scope of this paper. I propose briefly to discuss some consequences of the downturning of the paroccipital process and tabular, and to attempt to account for the conditions observed

in *Lacerta*, which may be taken as a fairly typical reptile so far as this region is concerned.

The downturn of the tabular and paroccipital process would involve a change in direction of the axis of the stapes and of the Eustachian tube; instead of running postero-dorso-laterally they would run postero-laterally, and since in most of these reptilian forms the quadrates are also displaced forwards, the direction would become almost lateral (cf. figure 5B). The result would be that the processus tympanicus would be driven down rather close to the processus quadratus of the stapes, and the tympanic diverticulum above the latter would become much compressed from above. But it was noted above that a second region for a possible extension of the hyomandibular pouch is just behind and above the angle of the jaw, and between the processus quadratus and the processus hyoideus of the stapes. It is possible that this extension was present in all very early tetrapods, but if so it was lost in the Stegocephalia. In the Reptilia it became important. I suggest that it thrust between the quadrate and ceratohyal connexions of the stapes, reached up toward the extrastapes (processus tympanicus) and also spread round the processus quadratus to meet the dorsal diverticulum (figure 3E, F). A similar process may be observed during the ontogeny of *Lacerta* (Goodrich 1915, especially plate 11, figures 6-10, and also Goodrich 1930, figure 477). By the union of these diverticula the stapes became enclosed in the tympanic cavity. The dorsal diverticulum of this account (which I shall call recessus tympanicus dorsalis, apparently homologous with the amphibian tympanic diverticulum) is the 'anterior inner recess of the tympanic cavity' of Goodrich (1915); the more ventral diverticulum (which may be called recessus tympanicus infrastapedialis), not represented in amphibians, divides around the extrastapes to form the 'anterior outer recess' and the 'posterior recess' during the ontogeny of *Lacerta*. In *Lacerta* the truncus hyomandibularis passes back between the processus dorsalis and the processus ventralis and then behind the stapes, while the chorda tympani passes over the extrastapes (processus tympanicus) and lateral to the connexion with the quadrate. In the diagram (figure 3C, E) it is shown that these conditions can readily be derived from those observed in Rhipidistia by imagining that the two rami separated more and more proximally, the hyoid ramus cutting through the stapes until free, the chorda tympani being freed from its tunnel perhaps by thinning of the whole stapes.

It is perhaps worth noting that failure of the dorsal tympanic recess ('anterior inner recess') would leave the chorda tympani in a pretympanic position; it might be possible better to understand the conditions in some birds on this basis.

In living reptiles and their fossil relatives it seems that the relations of the tympanic cavity to the hyoid arch elements can best be understood by comparison with the lacertilian conditions. But in the case of the theromorph reptiles, and perhaps in the case of some cotylosaurs, there seems to have been a further modification which is very significant from the point of view of the evolution of the mammalian inner ear.

6. THE MIDDLE EAR OF PELYCOSAURIA, THERAPSIDA AND MAMMALIA

Since the well-known work of Palmer, who compared the jaws of *Perameles* and cynodonts, it has been widely accepted that the tympanic annulus of mammals is homologous with the angular of the cynodont jaw, and in particular with that part of it forming a reflected lamina. A morphologically similar lamina occurs in all groups of 'Therapsida' and in the sphenacodont Pelycosauria which include their ancestors, and this was by many workers believed to have been the site of the tympanum in those animals. The clearest statement of this view is given by Sushkin (1927, 1928), who described the relief of the occiput and the posterior part of the mandible in *Captorhinus*, with details of pelycosaurs and other theromorphs, and suggested that there was a direct relationship between the posterior part of the mandible (and, in some forms, the region of the reflected lamina of the angular) and the tympanic cavity. The case of *Captorhinus* is not free from doubt; it seems possible that the groove on the inner face of the mandible may have transmitted the reptilian pterygoideus muscle.

In more recent years this view has been attacked, and there is now very clear evidence that the tympanum itself had no relationship to the angular flap in at least many forms of Theromorpha. This was shown most decisively by Broom (1936) in the cases of *Lycedops scholtzi* Br. and *Hofmeyria atavus* Br., therocephalians from the *Tapinocephalus* and *Endothiodon* (or *Cisticephalus*) zones respectively, where the stapes has a perfectly clear small extrastapedial process (= processus tympanicus) directed straight towards a partial circular opening on the occipital face, medial to the quadrate foramen, and bounded by the squamosal and quadrate, with the opisthotic almost entering the inner margin (figure 5D). Since then a similar extrastapedial has been found in other therocephalians, anomodonts, gorgonopsians, and in the peculiar form *Milleretta* (Broom 1938), which is doubtfully referred to the pelycosaurs (see also Watson 1942). A similar blunt process may be seen in Watson's figure of *Eriolacerta* (1931, figure 3). It is quite clear that in these forms the tympanum was not related to the mandible. Romer & Price (1940, pp. 62, 83) have shown that in pelycosaurs it is mechanically and morphologically impossible that the stapes could have had any functional relationship to the region of the angular notch, and have indicated the probable position of the tympanum not far from that expected by comparison with therocephalians. Their argument seems reasonable, particularly in view of the condition in *Lycedops* and *Hofmeyria* mentioned above, and of the position of the groove for the external auditory meatus on the squamosal in cynodonts and other 'therapsids', and in the remarkable Chinese mammal-like form *Bienotherium* (Watson 1942).

The function of the angular notch was not, therefore, to support the actual tympanum. Broom has long held that it housed a submaxillary gland. Romer & Price, working on pelycosaurs, adopted the suggestion that the notch was related to the great development of the pterygoideus musculature which was attached to the inner surface of the lamina. The mammalian condition does not follow naturally

from this, for in mammals the *m. tensor tympani*, the homologue of the reptilian *m. pterygoideus*, is attached, not to the tympanic ring, but to the internal face of the malleus.

It seems much more probable that in *Dimetrodon*, for example, the reptilian pseudotemporal muscle (i.e. the adductor mandibulae internus, pars pseudotemporalis of Brock's nomenclature—cf. Parrington & Westoll 1940, p. 340—or the levator mandibulae anterior of Edgeworth's interpretation) would have been inserted on the inner and lower surface of the bones surrounding Meckel's cartilage in this region, and that the reptilian pterygoideus muscle (the pterygoideus division of the adductor internus of Brock) would have passed back from the pterygoid flange to the special process on the articular lying antero-median to the articular surface and the retroarticular process. But I can see no good reason why the pseudotemporalis should have been inserted on the entire inner face of the angular lamina. It is more likely that it was applied to the lower surface of the bony box enclosing Meckel's cartilage, and it may have wrapped round this bar to reach up into the narrow dorsal part of the notch. During the further evolution of mammal-like forms the pseudotemporal mass must have migrated anteriorly to take up its mammalian insertion (as the pterygoideus internus) on the angular process of the dentary. This change would therefore presumably be correlated with the anterior movement of the angular 'flap' in cynodonts, and would later free the angular altogether from service as a muscle attachment.

In my opinion a pharyngeal diverticulum, either originally separate from the Eustachian tube and tympanic cavity, or much more probably an extension of the tympanic cavity itself (recessus mandibularis), lay at first internal to the lower part of the angular, and later was housed in the angular notch.* It would meet laterally the reptilian pterygoideus muscle, and also (dorsally and somewhat anteriorly) the lower part of the quadrate rami of the pterygoid and epipterygoid, which are greatly reduced in cynodonts and in bauriamorphs, the most mammal-like groups (cf. figure 6). Its development may be connected with the great depression of the roof of the normal reptilian cavity. It may possibly have acted also as a vocal sac or resonator (it shows some resemblance to the vocal sac of a frog), as thought by Dombrowsky and by Sushkin (1928, p. 278). Such a condition is easy to understand when the normal ontogeny of the mammalian tympanic cavity is considered (cf. Goodrich 1915), and it offers a very simple explanation of the relationship of the tympanic membrane to the homologues of the angular, articular, prearticular and quadrate in mammals (figure 6A-D). With the development of the new jaw-articulation in a plane lateral to the old one, the original tympanum would become very deeply placed. It is here suggested that during the shortening of the whole region (cf. Parrington & Westoll 1940) the part of the pharyngeal diverticulum related to the angular became closely adjacent to the original tympanic membrane,

* I find that Gregory (1929, p. 218, figures 112B, 113A) has also concluded that a tympanic diverticulum, though not the tympanic membrane itself, was associated with the notch in the angular, but he does not discuss the corollaries of this view.

and that the more superficial part of the lower cavity replaced the tympanum quite simply by gradual transfer of function. The approximation of the dorsal part of the tympanic diverticulum to the lower part of the cavity, related to the lower jaw, would be hindered so long as the quadrate-articular joint was still functional. When

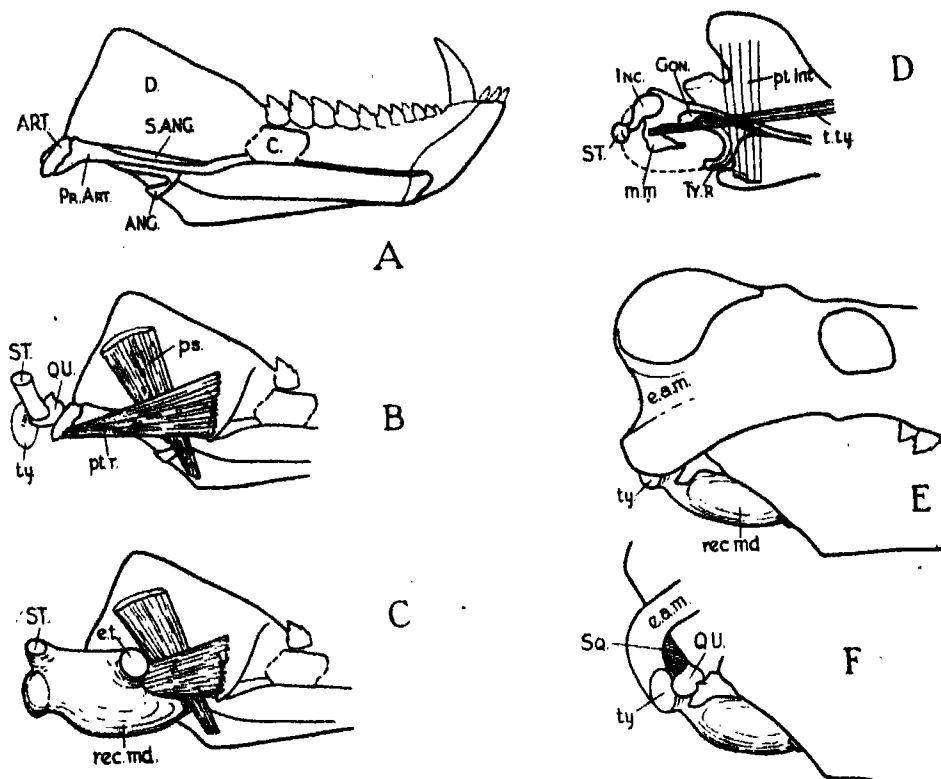


FIGURE 6. Suggested relations of the tympanic diverticulum in cynodont reptiles. A, inner view of left mandible of *Cynognathus* (after Broili & Schroeder). B, same with addition of quadrate, stapes, tympanic membrane, and the reptilian pterygoideus muscle (from pterygoid flange to inner side of articular) and pseudotemporal muscle (from alisphenoid to anterior part of angular and the angular process of dentary). C, same with addition of tympanic cavity showing processus ventralis of stapes (at fenestra ovalis), Eustachian tube, and recessus mandibularis housed in angular notch. D, inner view of mandible of pouch young of *Didelphys*, adapted from Goodrich (1930) with addition of the approximate courses of mm. tensor tympani and pterygoideus internus. The broken line suggests the extent of the tympanic membrane. E, external view of cheek and jaw of *Cynognathus* (adapted from Broili & Schroeder), showing suggested position of recessus mandibularis tympanicus. F, the same as E, with squamosal partly cut away to show probable course of external auditory meatus.

the dentary-squamosal articulation was established the tympanic diverticulum would be able to envelop the quadrate and the posterior end of the lower jaw, together with the posterior part of the remains of the reptilian pterygoideus muscle (homologous with m. tensor tympani), which probably passed from the pterygoid flange of the cynodonts and related theriodonts to the articular (cf. Parrington

& Westoll 1940, p. 341, and figure 14). The original tympanum (a thin membrane between the diverticulum and the hyomandibular cleft) could therefore extend downwards and forwards, and the more superficial part might be expected to take over the tympanic function. Then the retroarticular process and the possibly separate process (for the tensor tympani muscle) of the articular would naturally become embedded in the membrane, and would form the manubrium mallei, and the transformation of part of the m. pterygoideus into the mammalian tensor tympani also follows quite naturally. This suggestion explains simply and immediately why the angular should come to bear the mammalian tympanum, why the articular, prearticular and quadrate should become enclosed in the tympanic cavity, and why the stapes has lost its contact with the functional tympanum, and with it the extrastapedial process (which may, however, be represented in Paauw's cartilage). It explains also the difficulties, noted by several embryologists, that the tympanic cavity of mammals is morphologically somewhat antero-ventral to that of living reptiles, and that the attitude of the mammalian tympanum differs from that of reptiles.

Broom's view that the reflected lamina of the angular in many theromorphs housed the submaxillary (now more correctly called submandibular) gland requires comment. In many mammals this gland lies rather far forward, in front of the insertion of m. pterygoideus internus. But the parotid gland, which lies just behind the dentary below the condyloid process, immediately in front of and below the external auditory meatus, and below the level of Meckel's cartilage, occupies essentially the position of the upper and superficial part of the angular notch of advanced therapsids. It may be possible that both glands were primitively housed in the pocket covered by the reflected lamina, the duct of the ancestral parotid passing outside the coronoid process of the dentary and opening between the cheek and the teeth while that of the ancestral submandibular gland passed forward to open between the tongue and teeth in the usual way. Such a primitive condition would make it easy to understand the mammalian conditions, for expansion of the recessus tympanicus mandibularis into the notch might be expected to separate and displace the glands, the upper parotid part being pushed upwards and outwards, the lower moved forwards inside the dentary. But it is difficult to see good reasons for the general occurrence of such large salivary glands in reptiles like the sphenacodont Pelycosauria with bladelike or conical sectorial teeth. One might expect them to be present rather sporadically. I conclude tentatively that any relationship between these salivary glands and the pocket covered by the angular flap is likely to have been secondary, though it seems quite likely that the two glands were present in Cynodontia. The parotid gland was perhaps just below the exposed part of the articular (figure 6E, F) and the submandibular just within the angle of the dentary in front of the recessus mandibularia.

The exact function of the 'angular notch' in sphenacodont Pelycosauria is therefore difficult to decide. It had nothing to do with the reptilian pterygoideus muscle, but may have been occupied in part by the muscle (the reptilian pseudo-

temporalis) which became the mammalian pterygoideus internus. It may also (though in my opinion much less probably and not primarily) have housed a salivary gland or glands. Even at that early stage of the evolution of mammals I think it likely that a recessus mandibularis tympanicus had come into fairly close relationship with the region in question.

7. THE HOMOLOGY OF THE TYMPANIC CAVITIES OF TETRAPODS

Many workers have cast doubt on the strict homology of the tympanic cavity and membrane throughout the Tetrapoda. In particular, Gaupp (1913, especially pp. 294–306) stated, in his masterly work on Reichert's theory, that the tympanic membrane has such different relations in frogs, reptiles and mammals that it may have evolved independently from a generalized condition in which thick tissues separated the diverticulum from the outer skin. Sushkin (1927) also adopted the view that there are two positions of the whole tympanic diverticulum—an upper position above the jaw articulation, found in frogs and Stegocephalia and in many reptiles, and a lower position extending also beneath the jaw articulation, found in mammals and mammal-like reptiles. Goodrich, on the other hand, holds that the tympanic cavity and membrane are essentially homologous throughout (1930, pp. 478, 484–485). My own conclusions are somewhat different. They are that there was primitively one main diverticulum which, by pushing out new recesses may come to envelop different structures; that in labyrinthodont Stegocephalia and their allies and in living Anura the diverticulum was and is simple and very dorsally placed (recessus tympanicus dorsalis); that in many reptiles (including the relatives of living forms) there was a more ventral recess (recessus infra-stapedialis) from the diverticulum, arising between the quadrate and ceratohyal processes or ligaments of the stapes and enveloping parts of that complex; that in Theromorpha an additional recess (recessus mandibularis), still more ventrally directed, passed at first below the reptilian pterygoideus muscle and made contact with the posterior inner face of the mandible, being finally housed in part in the angular notch; and that in mammals the dorsal and ventral parts of the whole cavity of the theromorphs have enveloped the region of the old jaw articulation and that the tympanum has been able to enlarge or migrate antero-ventrally (see figure 3). In my view, therefore, the different conditions have arisen by evolution and modification of a primitive dorsal diverticulum which is essentially homologous throughout but is greatly reduced in mammals; it is the additional recesses that cause the difficulty. The modifications appear in each case to be correlated with the downward movement of the structures forming the roof of the tympanic cavity; unless otherwise compensated by new extensions of the cavity this process would have constricted the diverticulum very considerably, and perhaps obliterated it. These conclusions run more or less parallel to those of Gaupp, but the mechanism is quite different.

REFERENCES*

- Allis, E. P. 1922 Cranial anat. *Polypterus*. *J. Anat., Lond.*, 56, 189-204.
- Broom, R. 1936 The skull in Therocephalia. *Phil. Trans. B*, 226, 1-42.
- Broom, R. 1937 Fossil reptiles of the Karroo. *Proc. Zool. Soc. Lond.* for 1937, B, pp. 299-318.
- Broom, R. 1938 A primitive Upper Permian Reptile. *Proc. Zool. Soc. Lond.* B, 108, 535-542.
- Bryant, W. L. 1919 Structure of *Eusthenopteron*. *Bull. Buffalo Soc. Nat. Sci.* 13, 1-59.
- Bulman, O. M. B. & Whittard, W. F. 1926 On *Branchiosaurus*. *Proc. Zool. Soc. Lond.* for 1926, pp. 553-579.
- Bystrow, A. P. 1938 *Dvinosaurus*. *Acta Zool., Stockh.*, 19, 209-295.
- Bystrow, A. P. 1939 Blutgefässyst. Labyrinthodonten. *Acta Zool., Stockh.*, 20, 125-155.
- Bystrow, A. P. & Efremov, J. A. 1940 *Benthosuchus*. *Trav. Inst. paléozool. Acad. Sci. U.R.S.S.* 10 (1), 152 pp. (English pp. 120-150).
- Case, E. C. 1931 A new *Buettneria*. *Contr. Mus. Geol., Univ. Mich.*, 3, 187-206.
- Case, E. C. 1935 Skeletons of *Trimerorhachis*. *Contr. Mus. Geol., Univ. Mich.*, 4, 227-274.
- Eaton, T. H., jr. 1939a Evolution of the operculum in fishes. *Copeia*, pp. 42-46.
- Eaton, T. H., jr. 1939b The crossopterygian hyomandibular. *J. Wash. Acad. Sci.* 29, 109-117.
- Gaupp, E. 1913 Die Reichertsche Theorie. *Arch. Anat. Physiol., Supp. Anat.*, 1912, xiii + 1-416.
- Goodrich, E. S. 1915 The chorda tympani and middle ear. *Quart. J. Micr. Sci. (n.s.)*, 61, 137-160.
- Goodrich, E. S. 1930 *Studies on vertebrates*. 837 pp. London: Macmillan and Co.
- Gregory, W. K. 1929 *Our face from fish to man*. New York and London: G. P. Putnam's Sons.
- Parrington, F. R. & Westoll, T. S. 1940 The mammalian palate. *Phil. Trans. B*, 230, 305-355.
- Price, L. I. 1935 Brain case of *Captorhinus*. *Proc. Boston Soc. Nat. Hist.* 40, 377-386.
- Pusey, H. K. 1938 Changes in the anuran mandibular arch. *Quart. J. Micr. Sci.* 80, 479-552.
- Romer, A. S. 1937 Brain case of *Megalichthys*. *Bull. Mus. Comp. Zool. Harv.* 82, no. 1, 73 pp.
- Romer, A. S. 1939 Branchiosaurs. *Amer. J. Sci.* 237, 748-761.
- Romer, A. S. 1941 The crossopterygian hyomandibular and braincase. *J. Morph.* 69, 141-160.
- Romer, A. S. & Witter, R. V. 1942 *Edops*. *J. Geol.* 50, 925-960.
- Romer, A. S. & Price, L. I. 1940 Review of the Pelycosauria. *Spec. Pap. Geol. Soc. Amer.* no. 28.
- Säve-Söderbergh, G. 1936 Triassic Stegocephalians. *K. svenska VetenskAkad. Handl.* (3), 16, no. 1, 181 pp.
- Smith, J. L. B. 1939 A living coelacanthid fish. *Trans. Roy. Soc. S. Afr.* 28, 1-106.
- Sternberg, R. M. 1941 Cranial morph. of *Eusthenopteron*. *Univ. Toronto Stud., Geol.*, no. 45.
- Sushkin, P. P. 1927 The mandibular and hyoid arches in Tetrapoda. *Paleont. Z.* 8, 263-321.
- Sushkin, P. P. 1928 Cranial morph. of *Captorhinus*. *Palaeobiologica*, 1, 263-280.
- Watson, D. M. S. 1926 The Carboniferous Amphibia of Scotland. *Palaeontol. Hung.* 1, 221-252.
- Watson, D. M. S. 1931 Skeleton of a Bauriamorph reptile. *Proc. Zool. Soc. Lond.* for 1931, pp. 1163-1205.
- Watson, D. M. S. 1940 The origin of frogs. *Trans. Roy. Soc. Edinb.* 60, 195-231.
- Watson, D. M. S. 1942 On Permian and Triassic tetrapods. *Geol. Mag.* 79, 81-116.
- Westoll, T. S. 1942 Some primitive tetrapods. *Nature, Lond.*, 150, 121.
- Westoll, T. S. 1943 Origin of the tetrapods. *Biol. Rev.* 18, 78-98.
- Wilson, J. A. 1941 The skull of *Buettneria*. *Contr. Mus. Geol., Univ. Mich.*, 6, 71-111.

* Titles of papers have been abbreviated.

INDEX TO VOLUME 131 (B)

- Amino-acid analysis and the structure of proteins, Bakerian Lecture (Chibnall), 136.
Andrade, E. N. da C. Newton and the science of his age, 207.
Anniversary address (Dale), 191, 204.
- Badger, G. M., Cook, J. W., Hewett, C. L., Kennaway, E. L., Kennaway, N. M. & Martin, R. H. The production of cancer by pure hydrocarbons, VI, 170.
Bakerian Lecture. Amino-acid analysis and the structure of proteins (Chibnall), 136.
Barber, H. N. & Callan, H. G. The effects of cold and colchicine on mitosis in the newt, 258.
Bartlett, F. C. Ferrier Lecture. Fatigue following highly skilled work, 247.
Blood sedimentation (Whittington), 183.
Bristles in normal and mutant types of *Drosophila* (Lees & Waddington), 87.
Browning, H. C. The integument and moult cycle of *Tegenaria atrica* (Araneae), 65.
Bullough, W. S. The starling (*Sturnus vulgaris* L.) and foot-and-mouth disease, 1.
- Callan, H. G. See Barber & Callan.
Cancer, production by pure hydrocarbons, VI (Badger & others), 170.
Chibnall, A. C. Bakerian Lecture. Amino-acid analysis and the structure of proteins, 136.
Chromatic behaviour (Hogben), 111.
Compton, K. T. Pilgrim Trust Lecture. Organization of American scientists for the war, 297.
Cook, J. W. See Badger & others.
Croonian Lecture. Chromatic behaviour (Hogben), 111.
- Dale, Sir Henry Anniversary address, 191, 204.
Drosophila, bristles in normal and mutant types (Lees & Waddington), 87.
- Earthworm, visceral nervous system (Millott), 271, 362.
Evolution in plants by kaleidoscopic mutation (Willis), 161.
- Ferrier Lecture. Fatigue following highly skilled work (Bartlett), 247.
Fovea and extrafovea, spectral sensitivity (Walters & Wright), 340.
- Haemoglobin, fate in arthropods (Wigglesworth), 313.
Hewett, C. L. See Badger & others.
Hogben, L. Croonian Lecture. Chromatic behaviour, 111.
Hyomandibular of *Eusthenopteron* (Westoll), 393.
- Incompatibility in plants (Lewis), 13.
- Jeans, Sir James. Newton and the science of today, 231.
- Kennaway, E. L. See Badger & others.
Kennaway, N. M. See Badger & others.

- Lees, A. D. & Waddington, C. H. The development of the bristles in normal and some mutant types of *Drosophila melanogaster*, 87.
- Lewis, D. The physiology of incompatibility in plants. I. The effect of temperature, 13.
- Martin, R. H. See Badger & others.
- Mather, K. & Wigan, L. G. The selection of invisible mutations, 50.
- Millott, N. The visceral nervous system of the earthworm. I. Nerves controlling the tone of the alimentary canal, 271. II. Evidence of chemical transmission and the action of sympatheticomimetic and parasymphatheticomimetic drugs on the tone of the alimentary canal, 362.
- Mitosis in the newt, effects of cold and colchicine (Barber & Callan), 258.
- Newton tercentenary celebrations, 203-246:
 presidential address (Dale), 204;
 Newton and the science of his age (Andrade), 207;
 Newton as an experimenter (Rayleigh), 224;
 Newton and the science of today (Jeans), 231;
 exhibits, 243.
- Pilgrim Trust Lecture. Organization of American scientists for the war (Compton), 297.
- Rayleigh, Lord. Newton as an experimenter, 224.
- Rhodnius*, fate of haemoglobin (Wigglesworth), 313.
- Selection of invisible mutations (Mather & Wigan), 50.
- Starling and foot-and-mouth disease (Bullough), 1.
- Tegenaria*, integument and moult cycle (Browning), 65.
- Tetrapod limb, origin (Westoll), 373.
- Trichromatic theory of vision (Walters), 27.
- Waddington, C. H. See Lees & Waddington.
- Walters, H. V. Some experiments on the trichromatic theory of vision, 27.
- Walters, H. V. & Wright, W. D. The spectral sensitivity of the fovea and extrafovea in the Purkinje range, 340.
- Westoll, T. S. The origin of the primitive tetrapod limb, 373.
- Westoll, T. S. The hyomandibular of *Eusthenopteron* and the tetrapod middle ear, 393.
- Whittington, R. B. Blood sedimentation: a study in haemo-mechanics, 183.
- Wigan, L. G. See Mather & Wigan.
- Wigglesworth, V. B. The fate of haemoglobin in *Rhodnius prolixus* (Hemiptera) and other blood-sucking arthropods, 313.
- Willis, J. C. Evolution in plants by kaleidoscopic mutation, 161.
- Wright, W. D. See Walters & Wright.

I. A. R. I. 75.

IMPERIAL AGRICULTURAL RESEARCH
INSTITUTE LIBRARY
NEW DELHI.

[illegible]